Social Science
Library
FEB 25 1954

Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
HUMAN RESOURCES RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *Human Resources Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *Human Resources Research Center*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

Strength of Fear as a Function of the Number of Acquisition and Extinction Trials: H. I. KALISH 1

Some Effects of Prearranged Performance Scores upon the Level of Aspiration: J. R. SIMON, M. E. SHAW, AND J. C. GILCHRIST 10

The Familiarity-Frequency Relationship: C. E. NOBLE 13

Latent Learning in Children: H. W. STEVENSON 17

Identification of Dot Patterns from Memory as a Function of Complexity: R. S. FRENCH .. 22

Pattern Recognition in the Presence of Visual Noise: R. S. FRENCH 27

Effect of Shift in Distribution of Practice Conditions following Interpolated Rest: J. A. ADAMS AND B. REYNOLDS 32

The Contribution of Component Activities to the Total Psychomotor Task: E. A. BILODEAU AND I. MCD. BILODEAU 37

Postrest Performance in Motor Learning as a Function of Prerest Degree of Distribution of Practice: E. J. ARCHER 47

Primary Stimulus Generalization of the GSR under Two Conditions: D. D. WICKENS, H. M. SCHRODER, AND J. D. SNIDE 52

Induction Effects in Estimates of Extent: T. OBONAI 57

American Psychological Association

Vol. 47 No. 1 January 1954

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.50, or $15.00 annually. Single copies are $1.50. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1954 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 (4, Suppl.), 389–449. In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Copyright, 1954, by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 47, NO. 1 JANUARY, 1954

STRENGTH OF FEAR AS A FUNCTION OF THE NUMBER OF ACQUISITION AND EXTINCTION TRIALS[1]

HARRY I. KALISH[2]

State University of Iowa

Fear is usually defined as an anticipatory conditioned response to cues preceding pain (8, 10). It is assumed that fear functions as a cue (6) and as a drive (7). A number of studies (1, 2, 3, 6, 7, 9) have indicated that fear can motivate behavior and when reduced in strength can reinforce the immediately preceding responses.

In a recent survey of acquired drives, Miller (8) has indicated that if fear has the characteristics of a response, then its intensity, the probability of its occurrence, and its resistance to extinction should vary with those factors which affect the acquisition and extinction of other responses. On this basis, it is reasonable to expect fear to increase as a monotonic function of the number of conditioning trials and to weaken progressively with successive extinction trials.

The available data bearing on this expectation are fragmentary and inconclusive. In a study by Miller and Lawrence (9) no significant difference was observed between two groups of animals given 4 and 24 trials of fear conditioning. This failure to demonstrate the anticipated superiority of the 24-trial group was attributed, in part, to the fact that *Ss* in both groups received a large number of shocks during their preliminary trials. It was also suggested that fear may have reached a maximum in only four trials because the fear-arousing cues were very distinctive. The results of a study by Brown, Kalish, and Farber (2) support this latter suggestion since strongly conditioned fear reactions were noted after only three paired presentations of a CS and shock. It is conceivable, therefore, that although strength of fear may indeed be an increasing function of number of reinforcements, it may reach a maximum in substantially fewer trials than is commonly the case with conditioned skeletal responses.

In an investigation by Gwinn (3) designed in part to determine the relationship between number of reinforcements and strength of fear, one group of rats was given 4 and another 16 trials on each of which they were

[1] This report is a portion of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology, State University of Iowa. The writer is indebted to Dr. Judson S. Brown for his guidance during the experimental investigation and the preparation of the manuscript.

[2] Now at Duke University.

Social Science
Library
BF
1
.J86
VOL 47 1954

placed in a box and shocked. Intensity of shock was also varied in the same design. The results, measured in terms of persistence of escape from the box, indicated that resistance to extinction in the weak-shock group was greater for those animals given 16 fear-conditioning trials than for those given 4. The group given 16 strong shocks, however, showed less resistance to extinction than the group given 4 strong shocks.

The only available study with animals in which values have been obtained for more than two points on the function relating fear to number of trials is that of Brown, Kalish, and Farber (2). In this experiment tests for fear were interpolated during the course of fear acquisition, i.e., after varying numbers of acquisition trials. But, since shock was omitted from each test trial on which fear was measured, the procedure was essentially one of partial reinforcement. The function reflecting growth of fear under these circumstances may thus have suffered some distortion. One purpose of the present investigation, therefore, was *to determine the relation between strength of fear and number of acquisition trials independent of nonreinforced test trials.*

With respect to extinction of fear in animals, no studies exist in which fear has been measured after varying numbers of extinction trials, and in which an extinction trial is defined, in accordance with the classical conditioning paradigm, as the presentation of the fear-arousing CS in the absence of the UCS. In most of the aforementioned studies (1, 3, 7, 9) the CS was presented without shock and extinction did occur. But the duration of the CS could not be controlled because it depended on the animal's response. Thus, although fear may have been undergoing some extinction on each escape trial, the response being measured typically showed an initial increase in strength because of the reinforcement provided by fear reduction. The fact that new responses were learned may be said to indicate the presence of fear, but the magnitude of the response at any moment cannot be said to reflect degree of fear. Brown, Kalish, and Farber (2) have provided some systematic data relevant to the extinction of fear under conditions in which the duration of the CS is controlled and no new response is learned. However, the introduction of a startle stimulus (used as an index of fear) during the shockless trials may have had a disruptive effect upon the course of extinction. The second purpose of the present experiment, therefore, was *to investigate extinction of fear as a function of number of nonreinforced trials under conditions in which no overt response was reinforced.*

Method

Apparatus

Two sets of apparatus were used: one for acquisition and extinction, and the other for the hurdle-jumping phase of the experiment.

In general, the acquisition procedure involved paired presentations of the CS and the UCS while extinction consisted of the presentation of the CS in the absence of the UCS. The acquisition and extinction phases of the experiment were conducted in the same apparatus which consisted of four boxes each constructed of ¾-in. pine, 11 in. long, 6 in. wide, and 5 in. deep (outside dimensions). These boxes were arranged in a row; four boxes were used to permit the running of four *S*s simultaneously. Since most *S*s had been used in studies involving black-white discrimination, the boxes were painted gray to obviate the necessity of equating for color preference. All the boxes were fitted with lids constructed by tacking hardware cloth to a wooden frame. The grid floors of the boxes were 3/32-in. brass rods mounted in bakelite strips at intervals of 7/16 in. Simulated guillotine doors were constructed at one end to increase the similarity between these boxes and the box used during hurdle jumping. A buzzer and light, which served as the compound CS, were attached to the other ends of the four boxes. The buzzers, 24-v. DC relays modified to produce the sound of a

buzzer, were fastened to the lids and were energized by 24-v. AC. Measurements obtained using a General Radio Sound Level Meter Type 756 indicated a sound level of 73 db inside one of the boxes with all four buzzers energized simultaneously. The lights, 115-v., 7.5-w. frosted bulbs, were mounted in holes drilled into the end panels directly below the buzzers and protruded 5/16 in. into the boxes. The illumination in these boxes, provided by a shaded overhead light in the center of the room, was .9 ft.-candle as measured by a Macbeth illuminometer. When these lights were on, the illumination was increased to 5 ft.-candles. The light and buzzer were presented simultaneously and were interrupted three times per second in an effort to increase their distinctiveness. The UCS for the fear reaction consisted of a 60–70-v., 60-cycle shock delivered directly to the grids from a Variac connected to the 110-v., AC line. No current-limiting resistor was used since it was necessary to shock four *S*s simultaneously and since a commutator device, similar to that used by Skinner (12), was incorporated into the circuit to reduce the shorting effect of feces. The presentation of the CS and UCS was automatically controlled by electronic timers (4).

The hurdle-jumping phase of the experiment was conducted in a separate apparatus. Here the CS was presented without the UCS and *S*s were permitted to escape. The hurdle-jumping apparatus consisted of a box 22 in. long, 6 in. wide, and 5 in. deep (outside dimensions) of ¾-in. pine. The box was painted gray and divided into two equal compartments by a guillotine-type door which extended down through the top at the center and rested directly on the upper edge of a 2-in. partition. The two compartments formed by the guillotine door were fitted with lids. One of them contained a grid floor and a buzzer and light, and was identical with the boxes previously described. The other compartment was similar, but was without a buzzer and light and contained a smooth floor. The sound level in the hurdle-jumping box was approximately the same as that in the acquisition-extinction boxes. The latency of each hurdle-jumping response was measured to the nearest .01 sec. by means of a Standard Electric Timer. The timer started automatically when the door was raised (leaving a 2-in. barrier over which *S* had to cross) and stopped when the door was lowered by *E* following a response.

Subjects

The *S*s were 138 female, hooded rats from the colony maintained by the psychology department of the State University of Iowa. They ranged in age from 130–226 days at the beginning of the experiment. All *S*s had been used previously in studies involving form and color discrimination, but none had been exposed to shock. Prior to the experiment, they were fed 10 gm. of Purina Laboratory Chow per day for 10–15 days and were provided with a constant supply of water. Four days before the experiment proper, and during the course of the experiment, food and water were available in the living cages at all times.

Experimental Design

A 4 × 4 factorial design was employed in which the four columns consisted of the number of extinction trials: 0, 3, 9, and 27. The four rows of the design comprised the number of acquisition trials: 1, 3, 9, and 27. One group of 16 *S*s (one *S* assigned to each cell) constituted a replication and the entire experiment contained eight replications. A control group, consisting of 10 *S*s who received backward conditioning, was also incorporated into the design.

Acquisition of fear.—A single fear-acquisition trial consisted of the paired presentation of the compound CS and shock. A group of four *S*s, all receiving the same number of acquisition trials, was selected and its members placed in the individual acquisition-extinction boxes. In accordance with the experimental design, different numbers of trials, from 1 to 27, were given the separate groups. On each acquisition trial the duration of the CS was 5 sec., and 4 sec. after its onset the shock was presented for 1 sec.

In an attempt to extinguish the fear aroused by such cues as the box and grid equally for all groups, it was necessary to equate the amount of time each group was exposed to these cues. Accordingly, all *S*s were kept in the boxes for the same period of time (81 min.) and received the designated number of acquisition trials during this interim. At the end of the 81-min. period, one group of *S*s was returned to its living cages and another group obtained. The intertrial interval during acquisition and extinction was 3 min.

Extinction of fear.—Extinction was administered on the day following acquisition. During these trials the procedure was identical with that followed during acquisition except that no shock was presented. The *S*s were again run in groups of four.

Those *S*s assigned to the zero extinction group were placed in the boxes for 81 min., but were not presented with the CS.

The order in which the different groups of *S*s in one replication received treatment during acquisition and extinction was randomized throughout the eight replications.

Acquisition of hurdle jumping.—During the acquisition of hurdle jumping each *S* was run individually. Twenty seconds after having been

placed in the grid side of the hurdle box, *S* was presented with the CS and the sliding door raised. This enabled *S* to cross the hurdle and enter the second compartment after which the CS was terminated by *E*. The *S* remained in the compartment for a period of 10 sec. and was then removed to a carrying case in an adjoining room to await its next trial 4 to 5 min. later. During this interval three other *S*s were given one trial each. If *S* did not cross the hurdle within 60 sec., the door separating the two compartments was closed and a latency of 60 sec. was recorded for the trial.

All *S*s received 12 trials in the hurdle-jumping apparatus immediately following extinction. A minimum amount of time was allowed to elapse between extinction and hurdle jumping in an attempt to reduce spontaneous recovery of fear and its subsequent influence on the first-trial latencies. On the day following extinction and hurdle jumping, all *S*s received an additional 12 trials in the hurdle-jumping apparatus.

The acquisition and extinction treatments were evaluated in terms of their effect upon hurdle-jumping performance. The latency of the first trial in the hurdle-jumping apparatus was used as an index of level of fear following varying numbers of extinction trials for each level of acquisition. Since the first-trial latency is unaffected by the reinforcement provided by fear reduction, it should represent a relatively pure measure of the growth and decay of fear attributable to the acquisition and extinction trials. This is based on the assumption that the motivation due to fear will directly affect *S*s' latencies.

The latency of the subsequent trials in the hurdle-jumping situation should continue to reflect residual fear persisting from the acquisition and extinction trials. A decrease in these latencies would indicate the effect of fear reduction on the acquisition of hurdle jumping; increasing latencies obtained near the end of these trials would denote further extinction of fear.

Control group.—The ten *S*s in the control group were treated in the same way as those given 27 acquisition and zero extinction trials except that a backward conditioning procedure was employed in which the 1-sec. shock was followed after 15 sec. by the 5-sec. CS. This group was introduced for two purposes: (*a*) to ascertain whether performance in hurdle jumping was due to any inherently noxious characteristics of the CS; (*b*) to determine whether the cues of the fear-conditioning box were eliminated by keeping *S*s in the box for an 81-min. period following acquisition.

Results

First-trial latencies.—The latencies obtained from the first hurdle-jumping trials for the 16 groups were transformed to logarithms (plus a constant of 2.0 to avoid negative values) and summed over all extinction conditions for each level of acquisition. The resultant means for 1, 3, 9, and 27 fear-acquisition trials confirmed the theoretical expectation in showing a slight progressive decrease in latency as a function of increasing numbers of acquisition trials. Since the differences between these means were small and the variances large, an *F* test showed no reliable difference.

Latencies for Trials 1–24, Days 1 and 2.—The response latencies for each animal on Trials 1–24, Days 1 and 2, were summed over blocks of three trials and the means transformed as above to increase the stability of the measures and satisfy the assumptions involved in the statistical analyses. The results, organized in this manner, were treated by an analysis of variance in which the four levels of acquisition, four levels of extinction, and eight replications constituted the between–*S* comparisons, and the four blocks of three trials comprised the within–*S* comparisons (5). This analysis (Table 1) made it possible to examine simultaneously the performance curves of all 16 experimental groups and to test the hypothesis that these curves are parallel to each other, i.e., that their trends do not differ.[3]

[3] Bartlett's test for the Day 1 analysis revealed that the hypothesis of homogeneous variance was not tenable. Recent empirical evidence (11), however, has indicated that the disparity between the obtained and theoretical *F* is not marked despite extreme heterogeneity of variance. Nevertheless, the probability values for the Day 1 analysis were doubled in order to justify more completely the conclusions drawn from the data.

TABLE 1

SUMMARY OF ANALYSIS OF VARIANCE: DAY 1

Source	*df*	Mean Square	*F*
Between *Ss*	127	.57	
Acquisition	3	4.62	15.00*
Extinction	3	4.45	14.44*
Replications	7	.55	
A × E	9	.66	2.13
A × R	21	.26	0.84
E × R	21	.57	1.87
A × E × R	63	.31	
Within *Ss*	384	.07	
Trials	3	2.74	
A × T	9	.19	4.34*
E × T	9	.22	5.07*
R × T	21	.04	0.79
A × E × T	27	.03	0.80
A × R × T	63	.05	1.05
E × R × T	63	.05	1.05
A × E × R × T	189	.04	
Total	511		

* Significant at the 1% level.

The obtained *F* for the triple interaction (Acquisition × Extinction × Trials) failed to exceed unity and the hypothesis of parallel curves could not be rejected. In view of this, the double interactions, Acquisition × Trials and Extinction × Trials were examined. The mean values for the curves are plotted in Fig. 1 and 2.

An examination of Fig. 1 reveals that the theoretical expectations were supported. The curves for Day 1 show that an increase in the number of fear-conditioning trials produced an increase in the slope and a decrease in the over-all mean latency. Further evidence for this is revealed in the tests for interaction and vertical displacement (Table 1) which were significant beyond the 1% level of confidence. A further analysis of the trends for the individual curves discloses a decrease in latency for every level of fear acquisition since the *F* in each case was significant beyond the 1% level.

The results for Day 2, as plotted at the right of Fig. 1, were also analyzed in the above manner and show a continued, marked superiority of the 27-trial group, no apparent extinction for any group, and a shift in the relative position of the 9-trial group. The test of interaction failed to attain significance at the 5% level. The test for vertical displacement was significant beyond the 1% level of confidence.

The data illustrated in Fig. 2 (Day 1) indicate that performance in hurdle jumping is inversely related to the number of fear-extinction trials and suggest that the extinction conditions of the present experiment were successful in diminishing differentially the fear aroused by the CS. This is substantiated by the tests for the main effects of extinction and the Extinction × Trials interaction which were both significant beyond the 1% level. Tests of trend for the individual curves revealed a decrease in latency which was significant at the 1% level for Groups 0, 3, and 9. The *F* for Group 27, however, was not significant. A similar analysis of the curves for Day 2 (Fig. 2) revealed that the hypothesis of no-interaction could not be rejected with any significant degree of confidence. The over-all *F* for vertical displacement, however, was significant at the 1% level.

Analysis of simple trends for zero extinction level.—The values plotted in Fig. 3 were obtained from the groups given four levels of acquisition, but no extinction trials. The curve for the zero extinction group in Fig. 2 (Day 1) is a composite of these four curves. The control group was also included in this analysis since this group received 27 shock trials (backward conditioning) and no extinction trials. The extremely large interaction found for these curves ($F = 8.58$, $p < .01$) is due solely to the presence

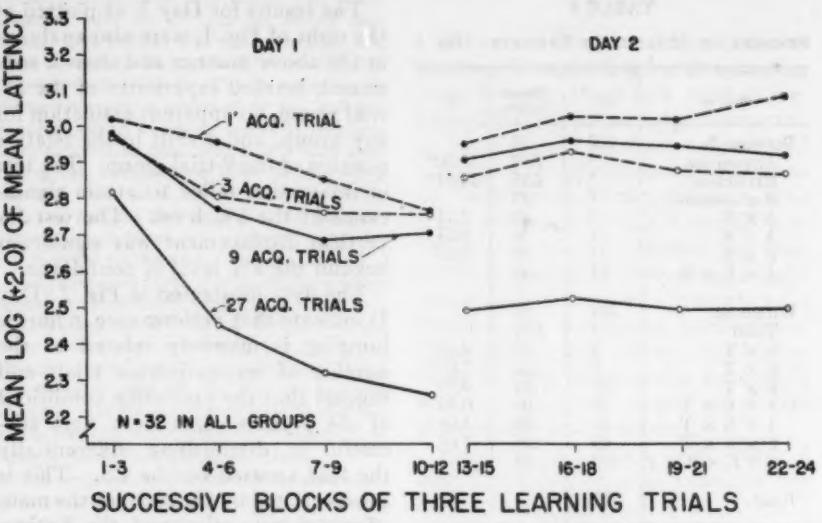

FIG. 1. Response latency as a function of successive hurdle-jumping trials for groups given varying numbers of fear-acquisition trials. All levels of extinction pooled for each level of acquisition.

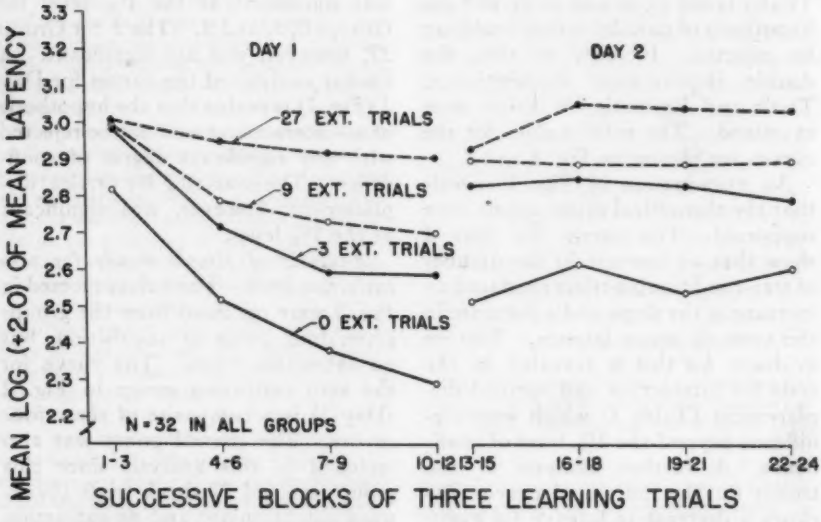

FIG. 2. Response latency as a function of successive hurdle-jumping trials for groups given varying numbers of fear-extinction trials. All levels of acquisition pooled for each level of extinction.

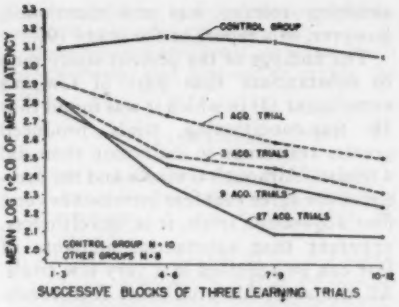

FIG. 3. Response latency as a function of successive hurdle-jumping trials for fear-acquisition groups 1, 3, 9, 27, and 27 (control) at the zero extinction level (Day 1)

of the control group since no significant interaction was found in a separate analysis for the other groups.

The failure of the control group to evince any decrease in latency with successive trials, as indicated by a test of trend, suggests that the CS was neither inherently noxious nor had acquired fear-arousing properties for this group. The performance of the control group also indicates that the procedure introduced to reduce the effectiveness of the box and grid as potential fear cues was successful, and that the significant (beyond 1% level) decrease in latency for the other groups can be ascribed to the acquired drive characteristics of the CS. The t tests for the separate means indicate that the control group differed significantly from the other groups beyond the 1% level. Of the remaining groups, the largest difference, between 1–0 and 27–0, was found to be significant at the 2% level for a single-tailed hypothesis.

DISCUSSION

The results of the present experiment provide additional support for the supposition that fear increases as a monotonic function of the number of fear-conditioning trials and weakens progressively with successive extinction trials. These conclusions are based primarily on the rewarding aspects of fear reduction rather than upon its motivating properties as revealed by latencies of the first hurdle-jumping trials. Although the reaction latencies of the four fear-acquisition groups ordered themselves in the anticipated manner on the first trial, these measures proved to be unsatisfactory as an index of level of fear because of their extreme variability.

Since, in the present instance, no crouching behavior was observed, this variability was due to the fact that *S*s differed widely in length of time taken to orient toward the guillotine door after being placed in the hurdle-jumping box. If controls were instituted so that all *S*s oriented more uniformly toward the second compartment, the measures obtained on the first trial might be more adequate indicators of level of fear.

The latencies of the subsequent trials in the hurdle-jumping box indicate that different numbers of acquisition and extinction trials resulted in significantly different levels of performance for the various groups (cf. Fig. 1 and 2). Within the framework of Hull's theory, there are two alternative explanations for these dissimilar performance rates and the resulting dissimilar asymptotic values reached by the various acquisition and extinction groups. Both of these alternatives are based on the assumption that performance is directly determined by reaction potential ($_sE_R$) which, in turn, is a function of the multiplicative combination of habit strength ($_sH_R$) and drive (D).

The first interpretation involves the assumption that the habit strength for hurdle jumping increased at the same rate for all groups as a function of successive reinforcements provided by fear reduction during the hurdle-jumping trials. The differences in performance reflected in the divergent curves would then be due to the unequal multiplicative effects of the different levels of fear-produced drive in the several groups. Thus, at the beginning of hurdle jumping the habit strength is low and generally equal

for all groups. Therefore, the multiplicative combination of ${}_sH_R$ and D would result in relatively small differences between the groups. If ${}_sH_R$ increased at the same rate for all groups, however, the habit strength at the end of the hurdle-jumping series would be high, but still very nearly equal for all groups. In this case multiplying ${}_sH_R$ and the differing D values would produce large differences between groups.

The second possibility is that the habits for hurdle jumping increase most rapidly for the groups having the greatest fear. According to this assumption the increments in ${}_sH_R$ per reinforcement would vary directly with the magnitude of the drive whose reduction provided the reinforcement. Although this does not seem to be a tenable assumption in the case of the effects of appetitive drives on habits, a recent eyelid-conditioning study by Spence (13) suggests that it may operate in the case where the drive is provided by noxious stimulation. In the present instance the increasing divergence of the curves in Fig. 1 and 2 would be attributable to both the differential rates of growth of ${}_sH_R$ and the multiplicative effects of the different drive levels.

It is interesting to compare the results of the present study with those obtained by Miller and Lawrence (9). These experimenters found that, although two groups of animals learned a wheel-turning response to escape fear-arousing cues after one group had been given 4 and the other 24 trials of fear conditioning, the difference between their levels of performance was not significant. In the present experiment, the two groups most nearly comparable to those in the Miller-Lawrence study were the groups given 3 and 27 fear-conditioning trials. Both of these groups showed a significant decrease in latency with successive hurdle-jumping trials, but the performance of the 27-trial group was significantly superior to that of the 3-trial group. Since strength of fear is also a function of the intensity of the UCS, it is conceivable that the failure of Miller and Lawrence to obtain a difference was due to the use of a more intense shock. The amount of shocking voltage was not mentioned, however, in a report of the study (9).

The findings of the present study tend to substantiate that part of Gwinn's experiment (3) in which it was found that 16 fear-conditioning trials produced greater resistance to extinction than did 4 trials. Although Gwinn's and the present study agree that fear increases beyond four acquisition trials, it is, nevertheless, apparent that substantial amounts of fear can be acquired in a very few trials. All the available studies of fear conditioning are in agreement on this point (2, 3, 9).

The failure of the control group to manifest any decrease in latency during hurdle jumping suggests that the CS was not in itself noxious and that the cues associated with the fear-conditioning box had not acquired any drive-producing characteristics. These findings support the notion that the performance of the other groups was due primarily to the acquired-drive characteristics of the CS.

Summary

The present experiment was designed to (*a*) determine the relation between strength of fear and varying numbers of acquisition trials, and (*b*) investigate the extinction of fear under conditions in which no new response is learned during extinction.

A total of 138 hooded, female rats, and a control group of 10 *S*s were run in a factorial study in which the experimental variables were (*a*) number of fear-acquisition trials (1, 3, 9, and 27), and (*b*) number of fear-extinction trials (0, 3, 9, and 27).

All *S*s went through experimental phases which involved acquisition of fear (paired presentations of CS and shock), extinction of fear (CS without shock), and acquisition of hurdle jumping. During the hurdle-jumping trials the CS was presented without shock and *S*s were permitted to jump a hurdle and enter a second compartment. After this response was completed the CS was terminated by E. The latency of hurdle-jumping response was measured. A backward conditioning procedure was employed for the control group to control for possible sensitization effects.

The results indicated that the experimental variables had a significant effect upon the acquisition of hurdle jumping. It was concluded that fear increases as a monotonic function of number

of fear-conditioning trials and weakens progressively with successive extinction trials.

Two alternative interpretations within Hull's learning theory were proposed to explain the results.

References

1. Brown, J. S., & Jacobs, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, **39**, 747–759.
2. Brown, J. S., Kalish, H. I., & Farber, I. E. Conditioned fear as revealed by magnitude of startle response to an auditory stimulus. *J. exp. Psychol.*, 1951, **41**, 317–328.
3. Gwinn, G. T. Resistance to extinction of learned fear-drives. *J. exp. Psychol.*, 1951, **42**, 6–12.
4. Hunter, T. A., & Brown, J. S. A decade-type electronic interval timer. *Amer. J. Psychol.*, 1949, **62**, 570–575.
5. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.
6. May, M. A. Experimentally acquired drives. *J. exp. Psychol.*, 1948, **38**, 66–77.
7. Miller, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.
8. Miller, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.
9. Miller, N. E., & Lawrence, D. H. Studies of fear as an acquirable drive: III. Effect of strength of electric shock as a primary drive and of number of trials with the primary drive on strength of fear. 1949. Quoted in (8).
10. Mowrer, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, **46**, 553–565.
11. Norton, D. W. An empirical investigation of the effects of non-normality and heterogeneity on the *F* distribution. Unpublished doctor's dissertation, State Univ. of Iowa, 1952.
12. Skinner, B. F. An automatic shocking grid apparatus for continuous use. *J. comp. Psychol.*, 1947, **40**, 305–307.
13. Spence, K. W. Learning and performance in eyelid conditioning as a function of the intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.

(Received April 6, 1953)

Journal of Experimental Psychology
Vol. 47, No. 1, 1954

SOME EFFECTS OF PREARRANGED PERFORMANCE SCORES UPON THE LEVEL OF ASPIRATION

J. RICHARD SIMON, MARVIN E. SHAW, AND J. C. GILCHRIST

University of Wisconsin

Previous performance in a particular situation has been found to be the most important variable affecting the level of aspiration (Lewin *et al.*, 2). Gardner (1), using a prearranged sequence of scores, found that goal discrepancies are positive and low when performance is improving and positive but markedly higher when performance is becoming progressively poorer. This finding was, however, incidental to Gardner's main problem, and, therefore, the experimental design was not such as to make the finding unequivocal. (*a*) Gardner used a single prearranged sequence of alternately increasing and decreasing scores and made no attempt to control possible overlapping effects on level of aspiration of these two opposing trends in performance. (*b*) All scores were reported in percentile values implying group standards which bring into operation reference scales beyond those of simple past performance. The present experiment controls these two difficulties by varying the patterns of past performance and using absolute numerical scores in a novel task.

METHOD

The apparatus used, a "bowling game," consisted of a table 5 ft. long, 2 ft. wide, and 2 ft. high. At one end and on top of this table there was a boxlike unit which served as a support for a counter and which housed a series of relays. The top of the table served as an alley leading to this boxlike construction. The alley entrance to the box was draped by a curtain. The *S* was told that there was a target in the box but that it was hidden by the curtain. The *S*'s task was to roll a small steel ball down the alley and try to rack up as high a score as possible by figuring out the location of "the moving target." The Veedor counter, which was clearly visible to both *S* and *E*, was connected in series with the relays and could be activated by *E* by means of a hidden switch under *E*'s desk. Thus, each time *S* rolled the ball, *E* could give *S* as high a score as desired and at the same time initiate a series of clicking noises which were intended to give the impression that the score was the result of the ball striking the hidden target. After each roll, the ball was automatically returned to *S*.

Eighty college students were used as *S*s. Each *S* was assigned randomly to one of four groups of 20 *S*s each. Each *S* was brought into the experimental room and seated in front of the apparatus. The *E* then said: "This is a complicated bowling game. It is complicated for two reasons. First of all, you can't see the target you are aiming at, and secondly, the target does not remain stationary. It moves from time to time. Your task is to roll the ball down the alley and rack up as high a score as you can. The highest possible score on any one ball is 20. There are five balls to a trial so you can make a maximum of 100 points on a trial. Again, this is a complicated game because you can't see the target and because the target moves from roll to roll, but there *is* a system to it, so that this is really a measure of your problem-solving ability. O.K., would you like to try it?"

After *S* had completed a five-roll trial, *E* asked him to read off the score, and pretended to record it. Then *E* asked, "What score do you really think you can get next time?"

The instructions and procedure were exactly the same for all groups except that Group I was presented with a prearranged sequence of attained scores starting at 31 and increasing 5 points with each block of two successive trials until a score of 71 was reached (the increasing series); Group II was presented with a prearranged sequence of attained scores starting at 71 and decreasing 5 points with each block of two successive trials until a score of 31 was reached (the decreasing series); Group III was presented with a prearranged sequence of scores starting at 31 and increasing by 5 points with each block of two successive trials until a score of 51 was reached, and then decreasing 5 points with each block of two successive trials until a score of 31 was reached (the increasing-decreasing series); Group IV was presented with a prearranged sequence of scores starting at 71

and decreasing by 5 points with each block of two successive trials until a score of 51 was reached, and then increasing by 5 points with each block of two successive trials until a score of 71 was reached (the decreasing-increasing series). Each group was given 18 trials.

Results

Inasmuch as each successive block of two trials yielded either an increase or a decrease of five points in performance it was possible to compare the goal discrepancies over any multiple of two trials. Because the interest herein lay in comparing goal discrepancies during improving and degenerating performances as a function of previous experiences of improving or degenerating performance, the goal discrepancies between groups on the last four trials were compared. Table 1 presents the means and *SD*'s of the groups. A summary of the analysis of variance for these data is given in Table 2. Particular notice should be paid to the interaction term which indicates that if the second half was that of increasing performance, the magnitude of the goal discrepancy was different depending upon whether the first half was an increasing or decreasing sequence. This statement has, of course, its counterpart for the second-half-decreasing sequence. A similar analysis of the first four trials demonstrated that the groups did not differ in their initial goal discrepancies.

TABLE 1

Mean Goal Discrepancies and *SD*'s for Each Prearranged Sequence of Scores

Pattern of Scores		Mean	*SD*
First Half	Second Half		
Increasing	Increasing	2.56	2.24
Decreasing	Increasing	.90	3.48
Decreasing	Decreasing	6.86	5.77
Increasing	Decreasing	4.91	3.86

TABLE 2

Analysis of Variance for Mean Goal Discrepancies on the Last Four Trials

Source of Variance	*df*	Mean Square	*F*
First-half condition	1	.41	—
Second half-condition	1	345.49	21.14**
First-× second-half conditions	1	65.25	3.99*
Residual	76	16.34	
Total	79		

* p = <.05.
** p = <.001.

Discussion

It is clear that the data support Gardner's finding that the usual positive goal discrepancies were greater in degenerating performance sequences than in improving performance sequences. The magnitude of these differences, however, appears to be conditional upon previous performance experiences. Examination of Table 1 shows that in each case an inconsistent previous trend reduced the size of the goal discrepancy. This finding makes it clear that not only must the trend of present performance be taken into account in level of aspiration phenomena, but the pattern of past performance must also be considered. Of particular interest in this regard are the absolute values of the mean goal discrepancies. Examination of Table 1 shows that the increasing-increasing group aspired to a greater amount of change than the decreasing-increasing group. Likewise, the decreasing-decreasing group aspired to a greater amount of change than the increasing-decreasing group. In other words, it appears that an inconsistent previous performance resulted in the level of aspiration being more closely tied to the performance score than when the previous performance had been consistent.

Summary

The present study was designed to test some of Gardner's observations on the effects of performance changes on the level of aspiration when a prearranged sequence of scores was used. Four

groups of 20 *S*s each were given 18 trials on a miniature bowling alley in which *E* had complete control over the scores that were presented to *S* although *S* believed that he was earning the scores by rolling a ball down the alley. Group I was given an increasing series of scores, Group II a decreasing series of scores, Group III an increasing-decreasing series of scores, and Group IV a decreasing-increasing series of scores.

The results agree with those of Gardner in that the mean goal discrepancy of an increasing sequence of scores was found to be significantly smaller than the mean goal discrepancy of a decreasing sequence of scores, but the size of these discrepancies is changed significantly if the individuals have had different kinds of performance sequences prior to the sequence now occurring. These results also agree with previous findings in that the mean goal discrepancy was positive regardless of the nature of the preceding experience.

References

1. Gardner, J. W. Level of aspiration in response to a prearranged sequence of scores. *J. exp. Psychol.*, 1939, 25, 601–621.
2. Lewin, K., Dembo, T., Festinger, L., & Sears, P. S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 333–378.

(Received May 4, 1953)

Journal of Experimental Psychology
Vol. 47, No. 1, 1954

THE FAMILIARITY–FREQUENCY RELATIONSHIP[1]

CLYDE E. NOBLE

Louisiana State University

In a previous publication the writer (6) hypothesized that the judged familiarity (f) of verbal stimuli is some monotonic function of their objective frequencies of occurrence (n). It was further assumed, by derivation from an analysis of meaning (m) described in (4), that verbal stimuli also acquire m value as some increasing function of n. These premises were then utilized in inferring a positive correlation between f and m, by virtue of their postulated co-correlation with n. Evidence consistent with this deduction was reported in (6), the empirical equation being

$$f = 6.04\ (1 - 10^{-.067n}) + .28, \quad (1)$$

having an index of correlation,[2] p, of .92.

Both of these premises are derived notions. The familiarity-frequency function is hypothetically positive by the operations defining the attribute f, since *S*s were required to rate the items in terms of frequency estimates. Similarly, the meaning-frequency relationship is implied by the fact that the m value of a word is contingent upon frequency of stimulation as well as upon the variety of associations formed. This latter dependency is also superficially consistent with the findings of Thorndike (9) and Zipf (10), although their data are based upon variates which are not identical with the concepts of m and n used by the writer.

Since, logically, the meaning-familiarity relationship might have taken any correlation between zero and unity, the strengths of the two linkage correlations must be calculated before undertaking further analysis. If our premises are valid, the form and stability of the functions should be experimentally determinable. It will be convenient to begin with $f = \Psi\ (n)$.

The purpose of this paper, then, is to evaluate the exact functional relationship between familiarity (f) and frequency of stimulation (n). To many readers the positive sign of the Ψ function may appear intuitively reasonable, yet intuition does not reveal its form. As a matter of fact, some have a priori predicted linearity, others a logarithmic curve, and still others a power function. The experimental objective may also be regarded as a search for those independent, manipulable variables which underlie the R–R relationship between familiarity and meaningfulness (6). By distilling from complex R–R laws their fundamental S–R components, behavior theorists hope to acquire principles of increasing generality and predictive precision.

METHOD

General.—The aim of this investigation was to present visually a set of relatively unfamiliar and meaningless verbal stimuli in varying frequencies of occurrence (n). Upon each presentation *S* was required to pronounce the item aloud, thus insuring adequate stimulation and

[1] The experimental work for this study was performed as part of the United States Air Force Human Resources Research and Development Program. The opinions or conclusions contained in this report are those of the author. They are not to be construed as reflecting the views or indorsement of the Department of the Air Force.

This research was conducted in the Perceptual and Motor Skills Research Laboratory, Human Resources Research Center. S/Sgt R. B. Card supervised its administration.

[2] For a discussion of p see Guilford (2, p. 317).

appropriate response. Later, graphic judgments of the familiarity (f) of each word were collected on a five-point rating schedule. After transforming these data to an interval scale by a method used previously (6), the dependent f values may be functionally related to the independent n values.

Apparatus.—The exposure device consisted of a 48-item, 2 × 2 in. automatic slide projector (Selectroslide). The presentation rate, controlled by Hunter-Brown electronic timers, averaged 3.77 sec. with a mean stimulus exposure duration of 2.33 sec. Sixteen items were chosen from the low end of the f scale (6), their obtained f values ranging from .00 to .76. The following ten n values were then assigned to the words at random: 0, 1, 2, 3, 4, 5, 10, 15, 20, 25. Owing to restrictions imposed by the fixed capacity of the slide magazine it was necessary to repeat certain frequencies. Accordingly, $n = 1$ was assigned twice and $n = 2$ six times. The 15 different experimental stimuli, excluding $n = 0$ which served as a control, were reproduced in block letters and prepared as slides. These were mounted in the projector in two invariant random sequences of 48 presentations each. In order to equalize the frequencies of individual items in the two magazines, the randomization was appropriately restricted. In addition, successive presentations of the same stimulus were limited to two.

Procedure.—The experimental design consisted of 10 treatments (frequencies), 96 counterbalanced orders of presentation, and 3 *Ss* per cell. This technique confounds the interaction of treatment and order effects with individual differences. The procedure for a given group of 96 *Ss* required each *S* to begin the exposure sequence with a different stimulus item, thus removing any cumulative or position effects from the resultant f value comparisons. Following the frequency phase of the experiment, *Ss* proceeded to rate the 16 items for familiarity by the method of single stimuli, as described in (6).

The *Ss* consisted of 288 basic airmen at Lackland Air Force Base, assigned randomly to the order conditions. The instructions were simply that a pronunciation test would be given, that each word was to be pronounced whenever it appeared, and that the score would consist of the number of correct pronunciations. No one was informed that he would be asked later for familiarity judgments.

Rest periods of 2 min. were given between the two sequences and prior to the administration of the Familiarity Schedule. The *Ss* were tested individually during both phases of the experiment. Category labels on the rating form were the same as those used earlier (6), with the definitions appropriately modified for the present purpose.

Results and Discussion

The experimental findings are graphed in Fig. 1. Judged familiarity (f) is plotted as a function of frequency of stimulation (n). The obtained x/σ f values were calculated by the method of graded dichotomies, rejecting proportions $\geq .98$ and $\leq .02$. A test of internal consistency indicated an average error of only .01 in reproducing the original proportions from the scale values. This measure of the reliability of the dependent variable is consistent with earlier findings (6).

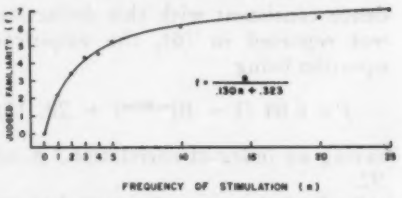

Fig. 1. The familiarity-frequency relationship

The f values in Fig. 1 have been adjusted by subtraction of the intrinsic *hypothetical* f values as derived from Equation 1. That is, the empirical median m value (4) of each stimulus item was substituted in the meaning-familiarity equation in order to calculate the pre-experimental "true" f values ($\bar{f}$). These $\bar{f}$ values were then used as correction factors to remove initial differences from the intertreatment comparisons. All values were finally readjusted to a zero origin, the ordinates corresponding to $n = 1$ and $n = 2$ having been averaged as noted in the procedure.

In Fig. 1 the familiarity-frequency function is observed to be hyperbolic in form. Its equation, computed by the method of least squares from a

linear plot of reciprocals, is

$$f = \frac{n}{.130\,n + .323}. \qquad (2)$$

The index of correlation p, calculated from the reduction line, has a value of .998. By successive differentiation it may be shown that the function has a positive slope and negative acceleration. The f asymptote has a value of 7.69. Thus we have sampled approximately 91% of the relative range of the dependent variable.

This empirical relationship between familiarity and frequency is obviously very stable and, therefore, of theoretical importance. As an example of a typical low-order S–R law, Equation 2 is now seen to clarify the common-sense notion as well as to rationalize the definition of familiarity proposed earlier (6). The present findings are thus consistent with the hypothesis that familiarity is a *learnable* attribute of stimuli. In terms of a general associational theory, Fig. 1 may be regarded as a performance curve determined by changes in some hypothetical learning process within the organism. The response variable f is thus considered as a magnitude function of physical events which are analogous to trials in a contiguity learning situation. As a consequence of these events, S's behavior is modified in a cumulative manner over trials (stimulus-pronunciation pairings). Although the intervening mechanism producing the familiarity judgments is unknown, the phenomenon is undoubtedly a memorial one.

It might be contended that Fig. 1 is not an acquisition curve at all, but that judgments of familiarity are uniquely determined by perceptual processes of comparison and discrimination whose dependency upon the physical variable n is nonassociational. That is, it might be argued that Equation 2 describes a *psychophysical* function of the classical type, and that the data follow a diminishing returns principle because differences among more frequently occurring stimuli are less discriminable than equal differences among less frequently occurring stimuli. The analogy is Fechnerian although the function is not.

Technically, neither the learning nor the perception formulation qualifies as a genuine explanation since each introduces a dispositional concept in the course of its analysis. They do, however, suggest different points of departure in subsequent theory construction. While learning psychologists would tend to theorize in terms of the *absolute* character of the responses in the rating situation, perception psychologists would prefer to attribute a relational or *comparative* property to the judgments. For example, Attneave's (1) recent study of the relationship between estimated and sampled relative frequencies of the letters of the alphabet is conceived within the framework of psychophysical theory. Disregarding their manifold differences, however, both of these investigations required S to evaluate antecedent physical conditions via present stimuli. This implies that some time-binding process like association is preliminary to absolute as well as comparative judgments in experiments not involving immediate report of sensory magnitudes.

A tachistoscopic study by Solomon and Postman (8) is relevant to this discussion. Using a training procedure similar to the present one, they found that visual recognition thresholds for "nonsense" (Turkish) words decreased as a curvilinear function of frequency of prior usage. Taken in conjunction with the curve in Fig. 1, this implies an inverse relationship between recognition thresholds and judged familiarity. Or, in terms of the reciprocals of exposure time, speed of recognition is positively correlated with familiarity. Further research in the field of perception should indicate whether, scalar considerations apart, these two measures are independent.

In connection with experiments in human learning, the results of this investigation indicate that the familiarity-frequency function, within the range tested, follows a hyperbolic law of diminishing returns. It serves, therefore, to distinguish clearly between physical and

psychological principles of "acquaintance" (7). This distinction is of heuristic value in suggesting possible systematic uses of response-defined concepts like *f* and *m* in behavior theory. For example, from the familiarity measure one might derive a mediating mechanism whereby the procedural variable *n* exerts its effects. This would constitute a rational source for postulating our intervening variables, instead of making arbitrary and often circular inferences from the phenomena to be explained. This technique would certainly not guarantee valid or useful concepts, but it might reduce the number of infertile hypotheses which theorists are likely to entertain.

One logical point of attack on this problem is to study further the nature of facilitation in verbal multiple-response learning. Specifically, what is the form of the effect on learning as *n* is varied? Some information is available (5) about the role of meaning (*m*) in learning verbal sequences, and an experiment by Hovland and Kurtz (3) has recently demonstrated the importance of familiarization in the same type of task. It is also known (6) that *m* and *f* are closely related, doubtless as a consequence of *n*. But the nature of the intervening processes associated with these observables is still in need of more comprehensive analysis.

Summary

Previous research had indicated that the judged familiarity (*f*) of verbal stimuli is some monotonic function of their objective frequencies of occurrence (*n*). In order to determine the exact form of this relationship, a sample of 16 words initially low in *f* value was presented randomly to 288 *Ss* in the following frequencies: 0, 1, 2, 3, 4, 5, 10, 15, 20, 25. Each item was pronounced aloud as it occurred.

Subsequently, *Ss* evaluated each word for familiarity on a five-point graphic rating schedule. These data were transformed to equal-interval *f* values, then adjusted for pre-experimental familiarity. When the adjusted *f* values were plotted as a function of *n*, a hyperbolic function resulted.

Rival interpretations of this empirical familiarity-frequency relationship were discussed.

References

1. Attneave, F. Psychological probability as a function of experienced frequency. *J. exp. Psychol.*, 1953, **46**, 81–86.
2. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
3. Hovland, C. I., & Kurtz, K. H. Experimental studies in rote-learning theory: X. Pre-learning syllable familiarization and the length-difficulty relationship. *J. exp. Psychol.*, 1952, **44**, 31–39.
4. Noble, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.
5. Noble, C. E. The role of stimulus meaning (*m*) in serial verbal learning. *J. exp. Psychol.*, 1952, **43**, 437–446.
6. Noble, C. E. The meaning-familiarity relationship. *Psychol. Rev.*, 1953, **60**, 89–98.
7. Robinson, E. S. *Association theory today.* New York: Century, 1932.
8. Solomon, R. L., & Postman, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43**, 195–201.
9. Thorndike, E. L. On the frequency of semantic changes in modern English. *J. gen. Psychol.*, 1948, **39**, 23–27.
10. Zipf, G. K. *Human behavior and the principle of least effort.* Cambridge, Mass.: Addison-Wesley, 1949.

(Received April 16, 1953)

Journal of Experimental Psychology
Vol. 47, No. 1, 1954

LATENT LEARNING IN CHILDREN [1]

HAROLD W. STEVENSON

The University of Texas

Latent-learning investigations with the white rat have been extensive; however, there have been few experiments on this topic with human *Ss*. A related topic, that of "incidental" learning in the human *S*, has been investigated, but the experiments involve a methodology quite different from that applied in the more recent latent learning experiments. It would seem desirable to translate the latent-learning problem to a situation applicable to human *Ss*, in this case, young children.

The present experiment is related to the type of latent-learning experiment (1) in which *S* is required to perform a certain task and is then tested to see if he has learned the location of objects encountered in performing the task, but which were irrelevant to his earlier motivation. Here, the child's task was to recover a reward from a locked box. To do this, he had to find a key in another box which contained, in addition to the key, several objects about which specific information was not needed to perform the task. After several trials, the child was requested to find one of the irrelevant objects.

The primary question is: Do children, motivated to find the key and reward, acquire specific information about the irrelevant objects, such as learning their location? Two further questions are: (*a*) whether the incidence of latent learning depends upon the chronological age of *S*, and (*b*) whether the spatial or perceptual relationship of the irrelevant objects and the key influences latent learning.

METHOD

Subjects.—The *Ss* were children residing in Stanford Village, a housing project of Stanford University, and a neighboring residential area. They were randomly divided into three groups of 45, and each *S* was tested under one of the three conditions to be described. On the basis of a pilot experiment, the age range from three to seven was chosen. In each condition there were 10 *Ss* of chronological ages four, five, and six, and 15 *Ss* of age three.

Materials.—A V-shaped "maze" was used, with a starting point located at the apex and goals located at the ends of the arms of the V. By placing the goals 10 ft. apart and 20 ft. from the starting point, a definite choice of path to each goal was necessary. On a small platform at each goal was a set of two small wooden boxes (8-in. cube). One box in each set of goal boxes was painted blue and the other yellow. Each blue box was locked with a large padlock and contained the reward object (a bird, flower, or animal sticker). Each yellow box was unlocked and contained the key and the irrelevant objects. One of the irrelevant objects in each yellow box was designated the test object: on the left side, a small purple matchbox and on the right side, a small white flannel purse. The other irrelevant objects were the same in each yellow box and were mere fillers, placed in the box to keep the test objects from being unduly conspicuous. They were nondescript objects with little association value (identical black and yellow cloth pads and similar hinges). The position of the objects in the box was the same on every trial.

Experimental conditions.—Three conditions were created by varying the spatial, or perceptual, relationship between the keys and the irrelevant objects. In Cond. 1, the key was on the test object, and could be obtained without manipulating or attending to the irrelevant objects. In Cond. 2, the key was under the irrelevant objects, and it was necessary for the child to manipulate the objects, if only in a cursory manner, in order to get the key. In Cond. 3, the key was in the test object, and thus the test

[1] Based upon a dissertation submitted to the Committee on Graduate Study, Stanford University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author expresses his gratitude to Drs. E. R. Hilgard, L. M. Stolz, and D. H. Lawrence for suggestions regarding this study.

object had to be picked out from the other objects and opened.

Procedure.—Each *S* was approached individually and given a large lock to play with. After *S* had mastered opening the lock, *E* suggested that he play a game with locks and took *S* out-of-doors to a large, protected area for the experiment.

A preliminary training period was introduced to aid *S* in learning how to open a locked box. A set of goal boxes was placed in front of *S*, the key was placed on the yellow one, and *S* was directed to take the key from the yellow box and open the lock of the blue box. This was repeated until *S* showed facility in performing this task. The *S*'s attention was then diverted by *E* while an assistant arranged the materials for one of the experimental conditions. The assistant then took a position midway and behind the two goals, and the training trials were begun.

The *E* directed *S* to the starting point and pointed to one set of goal boxes saying, "See those boxes down there? You find the key in the yellow box. Open the lock and see what is inside! All right?" If *S* did not appear to understand, or if he hesitated during the performance of the task, the instructions were repeated. Six trials were given each *S*, three with each set of goal boxes. On Trials 3 and 5 the *S*s were allowed a free choice of side, but the other trials were directed so as to insure equal experience with each set of goal boxes. Half of the *S*s were initially directed to the right and half to the left.

After *S* opened the box and obtained the reward, *E* asked him to bring the reward to the starting point and to paste it on a large placard placed there. As the child pasted, the assistant replaced the lock, key, and reward.

Following Trial 6 a test was made for latent learning. The *S* was told, "This time I want you to do something for me," and was shown a purse or box identical to the one in the yellow box. For half of the *S*s the purse was used as the test object, and for half the matchbox was used. The *S* was then asked, "See this purse (box)? Do you know where another one is just like this one? You find another one just like this."

Results

In order to establish the presence of latent learning we may consider the child successful if he goes to the side containing the test object requested, regardless of whether he returns with it or not. By chance, therefore, half of the children should be successful, unless biasing factors were operating in the test trials.

Two factors might operate to distort chance. First, one of the test objects may have been especially attractive, so that it would have been easier to locate than the other test object. A contingency coefficient of .04 reveals that the test object requested is not significantly correlated with success or failure on the test trial. Second, turning preferences may have interfered with chance success. However, of the 81 *S*s who chose the same side on both of their free trials and who might have developed a turning preference, 39 were asked for the object on their preferred side, and 42 for the object on their nonpreferred side; thus this factor induces no bias.

Eighty-seven *S*s went to the correct side, and thus are considered successful. Of these, 77 obtained the correct test object and returned with it to *E*. The other 10 chose the correct box, but did not return with the test object. Forty-two *S*s chose the incorrect set of boxes; however, 23 of these spontaneously corrected their error and eventually returned with the correct test object. We cannot determine the basis for their ultimate success, and must judge their performance according to their initial choice. Finally, there were 6 *S*s who indicated that they had no knowledge of where the object was, either by saying that they did not know and resisting persuasion, or by searching about the lawn and shrubs.

If we consider only those *S*s who approached one of the two sets of boxes as fulfilling the conditions of the experiment, the six who showed irrelevant behavior are thereby eliminated. For the 129 *S*s who did make a choice, 87 (67%) went to the correct box. Chance success was exceeded by

TABLE 1

PROPORTIONS OF *Ss* WHO CHOSE THE CORRECT GOAL BOX

CA	*N*	Condition		
		1	2	3
3	42	.73	.31	.64
4	29	.40	.30	.89*
5	28	.62	.60	.90*
6	30	.90*	.90*	.90*

* Significant at the .01 level.

17% ($p < .001$), and we may conclude that latent learning was shown.

In Table 1 the data are broken down by age and condition. A much higher proportion of success is seen in Cond. 3 than in either Cond. 1 or 2. Since the number in each group is so small, only the results for the 6-yr. group in Cond. 1 and 2, and the 4-, 5-, and 6-yr. groups in Cond. 3 are statistically significant ($p = .01$).

In order to determine whether the differences with chronological age and experimental conditions are significant, a more meaningful criterion of success than that of merely going to the correct side would be desirable. Complete success on the test trial required more than locating the test object, for *S* was also asked to return with it to *E*. By considering *S* successful only if he performs both of these acts, we may score the performance with somewhat more confidence. In using this criterion, the probability of meeting all these requirements by chance is less than 50%. However, we are now interested in comparing latent learning at different ages and under different conditions and this will not be of importance.

Figure 1 shows the proportion of *Ss* meeting the criterion of complete success at each age level. There is a general tendency for the proportion to increase at each successive age level. The significance of this increase was tested by a chi-square test. Since we are interested here primarily in the question of whether older children show latent learning more frequently than the younger children, we may combine the results for all three conditions and test whether the proportion of children in each age level increases, regardless of the condition under which the child was tested. The chi square obtained is 24.3 ($df = 3$), and is significant at the .001 level. The ability as measured and defined shows a very significant tendency to increase with chronological age.

In considering the differences in latent learning in the three training conditions, it can be seen from Fig. 1 that the proportion of success in Cond. 3 is considerably greater than in the other two conditions. Condition 1 provides the most inconsistent results, but in general seems to follow the pattern found in Cond. 2. When the scores of all *Ss* in each condition are combined, without regard for age, chi square is 10.58 ($df = 2$), which lies between the .01 and .001 levels of significance. Different conditions of

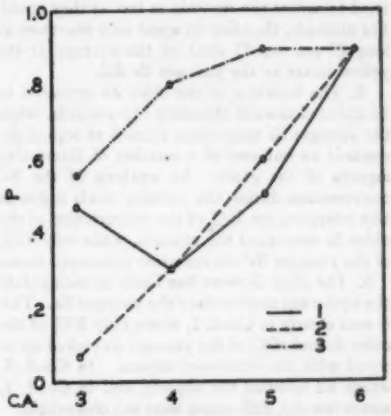

FIG. 1. The proportion of *Ss* in each condition and of each chronological age who chose the correct goal box and returned with the test object

stimulus relationship, therefore, have a significant effect on latent learning.

To test the pairs of conditions between which there is a significant difference, chi squares, each with 1 *df*, were computed. The *p* values for differences between Cond. 1 and 2, 1 and 3, and 2 and 3 were .30, .05, and .001, respectively.

Discussion

There are many functions associated with chronological age that might be related to the greater success of the older children. Certainly such factors as improvement in memory, spatial orientation, and perceptual abilities are of significance. Probably some of the most pertinent developmental changes are recorded in the observational records made during the experiment, and may be summarized as follows:

1. The older *S*s (ages 5 and 6) performed rather strictly according to the instructions given, while the younger *S*s (ages 3 and 4) appeared to look at the experiment as an opportunity for broad playful activity. The younger *S*s tended to play during their whole performance; they skipped and hopped to the boxes, sang as they opened the boxes, played with the locks, etc. The older *S*s ran to the boxes and tried to secure the rewards as fast as they could. For example, the older *S*s spent only one-third as long (5 sec. vs. 15 sec.) on the average at the yellow boxes as the younger *S*s did.

2. The behavior of the older *S*s appeared to be directed toward obtaining the rewards, while the younger *S*s more often seemed to regard the rewards as but one of a number of interesting aspects of the game. An analysis of the *S*s' conversation during the training trials supports this inference, for 56% of the conversation of the older *S*s concerned the rewards, while only 39% of the younger *S*s' conversation concerned them.

3. The older *S*s were less likely to manipulate the irrelevant objects than the younger *S*s. This is seen clearly in Cond. 2, where only 30% of the older *S*s and 80% of the younger *S*s picked up or toyed with the irrelevant objects. In Cond. 3, where all handled the objects, and in Cond. 1, where few did, differences were not observable.

4. There appeared to be a difference in the *S*s' ability to comply with *E*'s request to seek the test object, with the younger *S*s less able to change the game from that of looking for rewards to that of looking for a test object. The fact that 37% of the incorrect *S*s, mainly the younger *S*s, returned with the reward on the test trial, while only 6% of those who made the correct choice returned with the reward, supports such an inference.

The experimental conditions influenced the amount of learning that the younger *S*s showed, but had less effect as chronological age increased. For all of the children, however, the placement of the key determined to a great degree the type of response that would be made to the irrelevant objects. In Cond. 1, only 6 of the 45 *S*s stopped to play with the irrelevant objects, while in Cond. 2, 26 of the 45 *S*s played with the objects. Whatever response *S* made to the objects in Cond. 1, then, tended to be visual, but in Cond. 2, when *S* was forced to make some motor response in his search for the key, he tended to manipulate the objects, or simply to brush all of the objects aside. In Cond. 3, everyone was forced to make specific responses to the test object on each trial, and each *S* had to discriminate it from the other objects in the box, remove and open it.

Putting all this together, we find several factors that seem to be conducive to the type of learning investigated. The developmental changes indicate that the older *S*s performed in a more consistent, organized fashion. Their compliance with *E*'s requests, their type of play, and their interest in the rewards gave their performance a consistency and goal-directedness that was missing in that of many of the younger *S*s. The differences between conditions indicate that learning was most satisfactory when the irrelevant object was directly integrated into the sequence of behavior and had to be responded to differentially during the training trials. Learning was poorer when the object was present in *S*'s visual field but bore little relationship to the task he was performing, and when the object functioned only as part of a barrier that could be removed in a haphazard fashion.

In summary, the experiment illustrates the fact that although latent learning may not appear with *S*s at one level of

maturity, it may be found in *S*s at a later level, and it indicates that subtle differences in the type and amount of experience that *S* has with objects irrelevant to his momentary needs may influence the degree to which he will acquire information about them.

Summary

Three groups of 45 children of ages 3 to 6 were motivated to find a reward in a box which could be opened only by a key hidden in a second, unlocked box. In the second box were several irrelevant objects, and the placement of the key among them differed for each of the three groups. For one group the key was found in one of the objects (test object); for another it was found under the object; and for a third on the object. Two sets of goal boxes, separated spatially, and containing different test objects, were used. After three visits to each set of boxes, *S* was asked to find one of the test objects.

The results may be summarized as follows:

1. The ability to locate the test objects differed significantly from chance, and latent learning was shown.
2. The ability to locate the object increased significantly with increasing chronological age.
3. The frequency of latent learning was significantly greater when the key was in the object than in either of the other arrangements. The "under" and "on" groups did not differ significantly from each other.

References

1. Spence, K. W., Bergmann, G., & Lippitt, R. R. A study of simple learning under irrelevant motivational-reward conditions. *J. exp. Psychol.*, 1950, **40**, 539–542.

(Received May 18, 1953)

Journal of Experimental Psychology
Vol. 47, No. 1, 1954

IDENTIFICATION OF DOT PATTERNS FROM MEMORY AS A FUNCTION OF COMPLEXITY [1]

ROBERT STANTON FRENCH

Perceptual and Motor Skills Research Laboratory, Human Resources Research Center

Ease of pattern identification from memory might be expected to increase with an increase in pattern complexity. It is conceivable, however, that patterns may become too complex to be identified easily, at least under certain conditions. The present study investigates this question of the relation of pattern complexity to identification from memory, using random patterns of dots as stimuli. Specifically, the number of dots for independent groups of *Ss* was varied from 1 to 12; within a given group of *Ss*, however, the task was to learn names for 12 random patterns all of which contained the same number of dots. Thus, the task was one of differentiating patterns on the basis of the placement of the dots within each pattern.

The choice of stimulus materials in the present case was influenced by some previous work on the discrimination of dot patterns (**3, 4**). These earlier studies identified some of the relevant variables in the task of discriminating between *pairs* of dot patterns. The number of dots in the pattern, the average separation of the dots, and the relative angular orientation of the two patterns compared were found to be important considerations. With certain principles already established, there appeared to be some advantage in utilizing the same apparatus and stimuli similar to those already investigated.

Patterns generated randomly would seem to be rather well suited to a study of the effect of complexity upon ease of identification. The problem here is one of establishing a scale of complexity with some semblance of unidimensionality. Using random patterns, we may define complexity in terms of the number of dots or elements in the pattern; but even here, more than one dimension is involved since one cannot vary number of elements without also varying pattern. This lack of independence of pattern and number means that a relatively large sample of patterns of each number of dots should be used, if the variance attributable to differences in pattern difficulty is to be averaged out.

It is apparent that increasing the number of elements in a pattern increases the number of ways the patterns may differ, and thus, the possible number of different patterns. There is a line of evidence, however, e.g., (**5**), which indicates that when the number of elements in a complex exceeds about eight, an individual can no longer perceive the pattern as a "whole" in a brief glance (and report number accurately, for example). It may be that for the task of identifying a group of patterns from memory a degree of detail beyond about eight elements is not only unnecessary, but possibly detrimental. Assuming such a point of "optimum detail" was found, the length of the original exposure interval, the retention interval, and the number of patterns to be differentiated are some of the parameters which might be investigated further; the present study has the modest purpose, however, of determining whether a "point of optimum detail" may be demonstrated at all in the task of identifying random patterns of dots.

[1] The data reported in this study were collected as part of the United States Air Force Human Resources Research and Development Program. The opinions or conclusions contained in this report are those of the author. They are not to be construed as reflecting the views or indorsement of the Department of the Air Force.

METHOD

Apparatus.—The Pattern Discrimination Apparatus described in previous reports (3, 4) was used. This device permits pairs or individual series of dot patterns appearing as pinpoints of light against a homogeneous background to be presented successively, with control of all relevant time intervals. In the present study *Ss*, in groups of eight, sat about 15 ft. in front of the apparatus and viewed a series of such dot patterns presented individually.

Patterns.—A series of 48 different dot patterns was constructed for each number of dots from 1 to 12. The population of possible spatial positions was represented by 112 light sources arranged in rows and columns 1 in. apart and forming a matrix of roughly circular outline. The dot patterns for the present study were constructed using a table of random numbers to insure that each of the 112 spatial positions had an equal opportunity of being represented in each pattern for a given number of dots. Thus the patterns were a random sample from the population defined by the limits of the matrix.

Procedure.—A modified paired-associates procedure was employed. Twelve patterns of a given number of dots were presented individually to a group of eight *Ss*. The *Ss* were told that each pattern represented the outstanding features of some well-known American city as it might appear from the air on a radar scope. They were told that their task was to learn the city represented by each pattern. Immediately before and after each pattern appeared, *E* called out the corresponding city name. The *Ss* were instructed to observe the patterns carefully and to try to associate each with its name. The *Ss* were not permitted to write anything down during this training series. Following the initial presentation, the patterns were presented again in a different order as a test series. Each *S* was provided with an answer sheet on which he was to write in sequence the first letter of the city name which he believed corresponded to each pattern. Each *S* was also provided with a slip of paper listing in random order all of the city names used. This procedure of a training series followed by a test series was repeated with a new random permutation of the patterns for each presentation. The experimental procedure consisted, then, of two training and two test series presented alternately. Each pattern was exposed for .5 sec. The interval between successive exposures was 5 sec. The test series followed the training series after the short interval required to give necessary instructions to the *Ss*.

Design.—The basic set of 48 patterns for each number of dots from 1 to 12 was arbitrarily subdivided into four sets of 12 patterns each. Independent groups of 16 *Ss* were tested with each of the sets. Thus, data were obtained from independent groups of 64 *Ss* for each number of dots from 1 to 12. This arrangement permitted a simple analysis of variance testing the mean square between dot conditions against the mean square for *Ss*.

Subjects.—The *Ss* were 768 basic trainees at Lackland Air Force Base. These men were an unselected sample from the pool of *Ss* available to the laboratory for testing. None had served previously in a similar experiment.

RESULTS

The primary data of the experiment are presented in Fig. 1. Each point on this graph represents the pooled mean percentage of error for 64 *Ss*. Mean percentage of error was obtained by summing the error scores for all *Ss* represented at a given point, dividing by the total possible error (no correction for guessing), and converting this ratio to a percentage. For a visual estimate of significance between successive points on the curve, dotted lines representing $M \pm 1.96\ \sigma_m$ have been included. It may be seen from Fig. 1 that mean percentage of error decreases up to the seven-dot condition and then increases again irregularly beyond this point, thus appar-

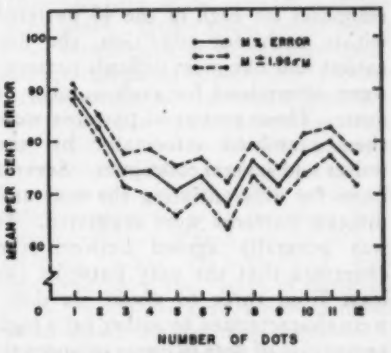

FIG. 1. Mean percentage of error in pattern identification as a function of the number of dots in the pattern. Each point represents data from an independent group of 64 *Ss*. Dotted lines representing $M \pm 1.96\sigma_m$ are indicated.

ently confirming the hypothesis under test. In view of the irregularity of the function, however, it was desirable to employ a more exact test of significance. It was obvious from an initial inspection of the data that marked inhomogeneity of variance was present. This was confirmed by an application of Bartlett's test (1). Accordingly, instead of an over-all analysis, *t*-test comparisons were made using *t'* as an approximation for *t* (2, p. 92).

Briefly stated, the differences between the means of the 7- and 3-dot conditions and the 7- and 12-dot conditions are both significant at the .05 confidence level. On the other hand, the differences between the 7- and 5- and the 7- and 9-dot conditions do not achieve significance at this level. From a conservative point of view, it may be said that the point of optimal differentiation probably lies between the 5- and 9-dot condition and most likely at 7 dots.

It was hoped that measurements of "pattern" variables might be related to ease of differentiation. The procedure used to carry out these measurements was as follows: from data indicating the total number of correct responses for each of the 48 patterns within each dot condition, the five easiest and five most difficult patterns were determined for each number of dots. These groups of patterns were then examined informally by the writer and several colleagues. Several bases for differentiating the easy and difficult patterns were suggested. It was generally agreed between the observers that the easy patterns (at least from three to about six dots) were characterized by either (*a*) a high proportion of dots in linear or approximately linear arrays, or (*b*) a high degree of symmetry.

As a measure of linearity the following procedure was employed: the average length of the perpendicular from each dot to a straight line drawn between the two nearest dots was obtained. Since it was desired to make the measure of linearity independent of the absolute distance between dots, the mean linearity value was divided by the corresponding mean distance of each dot to the two nearest, for each pattern. This linearity-ratio value was determined for each of the five easy and five difficult patterns for each number of dots. These ratio values are plotted in Fig. 2. In interpreting this plot it should be remembered that a low linearity-ratio value represents a pattern with high linearity. An examination of Fig. 2 reveals that high linearity does characterize at least one group of easy patterns. In the group of patterns from *three* to *six* dots, a clear separation is seen between the difficult patterns and *one* group of easy patterns (the lone exception in the three-dot group is a highly linear pattern which is almost identical to another in the same group). On the other hand, the linearity ratio does not distinguish, even for the three- to six-dot conditions, between the diffi-

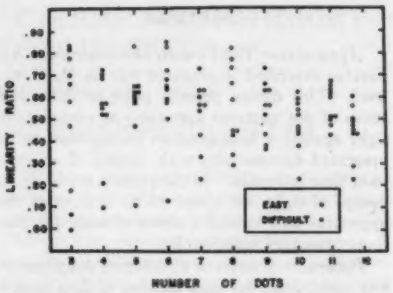

FIG. 2. A plot of the linearity-ratio values for the five easiest and five most difficult patterns for each number of dots. The ratio represents the average length of the perpendicular from each dot to a straight line drawn between the two nearest dots, divided by the mean distance from each dot to the two nearest. The ratio was determined for each pattern individually.

cult patterns and *another* group of easy patterns. This latter group of patterns was examined to discover why part of the patterns was among the five hardest and part among the five easiest. From even a casual examination it was clear that degree of symmetry was the basis for differentiation, in this case. The essential characteristic of symmetry, at least for dot patterns, would seem to be the duplication of corresponding angles (or distances) between dots on either side of a central axis. A rough index of symmetry was obtained in the following manner: measurement was made of the angle formed by lines drawn from each dot to the two dots nearest to it, taking each dot in the pattern in turn as the vertex. Thus, there were as many angles as there were dots in the pattern. Since a completely symmetrical pattern of three or more dots must have *at least* two equal angles when measured in this fashion, it was believed likely that a separation between the easy and difficult patterns might be made in terms of the *number* of equal or approximately equal angles in each pattern. This symmetry measure was applied for that group of patterns from three to six dots with linearity-ratio values above .40 (see Fig. 1). It will be noted that there are 14 easy patterns and 19 difficult patterns in the group so specified. Among the group of 14 easy patterns there were 13 cases of angles which were within four degrees or less of each other (one pattern had two sets of equal angles and two others had none). Among the group of 19 difficult patterns only 5 such cases of approximately equal angles were found, all of them in the six-dot patterns. The data clearly support the casual observation, therefore, that in patterns of from three to five dots a relatively high degree of linearity or symmetry characterizes the patterns that are most easily differentiated. Beyond six or seven dots the two factors (linearity and symmetry) are not so clearly delineated, although the easy patterns are, in general, characterized by relatively high linearity (Fig. 1).

Discussion

For the task of identifying dot patterns from memory a degree of complexity represented by six to eight dots was found to be optimal. Without further experimentation one can only speculate as to the generality of this finding. One might expect, for example, that if an individual were permitted a longer time to examine a pattern he could retain a more detailed impression; it is not at all clear, however, that this would be an advantage *unless* the number of patterns to be differentiated was quite large and the differences between the patterns slight. It should be kept in mind that any one group of *S*s was required to learn the names for only 12 randomly selected patterns out of a population many times this number (there are, for example, 227,920 possible different three-dot patterns in the matrix of 112 light points). Relatively speaking, the number of patterns to be differentiated here is exceedingly small and the differences between the patterns are likely to be gross. Under these conditions it seems rather unlikely that a longer exposure interval would do more than improve performance for all degrees of complexity. The hypothesis may be entertained that the difficulty in identifying patterns under the conditions of the present study lies not so much in the possible confusions between patterns, as in the sheer difficulty of associating each pattern with its arbitrary name. In accordance with this hypothesis, the "point of optimum detail" represents that degree of detail which maximizes the meaningfulness of the pattern. Perhaps there is a degree of complexity on the average at which a pattern is most likely to "look like something" or bring to mind useful associations. Casual observation of the sets of

patterns does tend to support this hypothesis; patterns composed of more than about nine dots seem cluttered and difficult to organize. The hypothesis would seem to be one deserving further exploration.

Summary

The identification of dot patterns from memory was studied as a function of the complexity of the patterns. Each of 768 basic trainees was required to learn the corresponding arbitrary city names for a group of 12 random dot patterns representing the possible appearance of each city on a radar scope. The number of dots in the patterns was varied from 1 to 12 for independent groups of *Ss* as a means of varying complexity.

The results indicate that a degree of complexity represented by six to eight dots is optimal for identification under these conditions. Ease of identification is clearly shown to be associated with patterns having dots arranged either symmetrically or in linear arrays.

References

1. Bartlett, M. S. Some examples of statistical methods of research in agriculture and applied biology. *J. R. statist. Soc. Suppl.*, 1937, **4**, 137–170.
2. Cochran, W. G., & Cox, G. M. *Experimental designs.* New York: Wiley, 1950.
3. French, R. S. The discrimination of dot patterns as a function of number and average separation of dots. *J. exp. Psychol.*, 1953, **46**, 1–9.
4. French, R. S. The accuracy of discrimination of dot patterns as a function of angular orientation of the stimuli. USAF Hum. Resour. Res. Cent., Res. Bull., 1953, No. 53–3.
5. Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. The discrimination of visual number. *Amer. J. Psychol.*, 1949, **62**, 498–525.

(Received May 20, 1953)

Journal of Experimental Psychology
Vol. 47, No. 1, 1954

PATTERN RECOGNITION IN THE PRESENCE OF VISUAL NOISE [1]

ROBERT STANTON FRENCH

Perceptual and Motor Skills Research Laboratory, Human Resources Research Center

The present study deals with the effect of visual "noise," i.e., randomly scattered points of light, on the recognition of random dot patterns of varying complexity. The task required *S* to recognize or "find" in a pattern of noise each of a series of patterns which appeared on a display screen. For convenience the stimulus patterns to be located in the noisy display were called the "target" patterns, while the same patterns embedded in the noise were referred to as the "scope" patterns or displays. The target and scope displays appeared side by side on the screen, enabling *S* to look freely from one to the other. To make the discovery of the target more difficult and to provide a means of measuring recognition, the scope display as a whole was rotated either to the right or left. Specifically, *S*'s task was to determine the direction of rotation; this could be done only if the target pattern was discovered in the scope display.

It was reasonable to suppose that increasing the complexity of the pattern to be recognized against a background of visual noise would increase the ease of recognition. As the target pattern was made more complex, more of the elements (dots) of the scope pattern would be, in terms of their spatial relations, identical to those in the target pattern alone. This increase in the number of common elements should increase the probability that some elements of the target would be recognized, even though differently oriented on the scope. On the other hand, if the amount of noise on the scope was increased for a given target, the increase in the total number of elements in the display should decrease the probability that target elements would be recognized. In other words, target recognition should improve with an increase in target complexity and become worse with an increase in visual noise. Defining complexity in terms of number of dots, the present study tested the hypothesis described above, varying the number of target dots from two to nine and the number of noise dots from one to eight. Although recognition performance was expected to improve with an increase in target complexity, previous work with the same apparatus on the *discrimination* (2) and the *identification* (4) of dot patterns suggested that for a relatively brief exposure (5 sec.), little further improvement in recognition might be found beyond six to eight dots. It was further hypothesized, therefore, that the function relating target recognition to complexity should be negatively accelerated, reaching an asymptote in the neighborhood of six to eight dots.

Although the effect of masking or camouflaging patterns by various means has been studied in detail, most of the work has dealt with qualitative variables; little work has been done on the quantitative aspects of the problem. One of the closest approxi-

[1] The data reported in this study were collected as part of the United States Air Force Human Resources Research and Development Program. The opinions or conclusions contained in this report are those of the author. They are not to be construed as reflecting the views or indorsement of the Department of the Air Force.

mations to the present task is the Camouflaged Figures Test (5) developed for the Aviation Cadet Classification Program. In this test, letters of the alphabet drawn as series of dots are embedded in a pattern of many irregularly spaced dots. Since the test has been used as a selection device, rather than as an experimental procedure, no data are available that might be related to the present study.

An interesting parallel might be drawn between the present approach and the concept of the "speech-to-noise" ratio developed primarily at the Bell Telephone Laboratories (1). The speech-to-noise ratio is the ratio in decibels of the speech intensity to the noise intensity. The concept has been useful in connection with computational procedures for the prediction of the intelligibility of speech under various noise conditions.

Using number of dots as the basis for a scale, one might determine recognition performance for various "target-to-noise" ratios. Such an approach could provide an approximation of the minimum ratio for efficient target recognition; such a ratio value might have considerable generality.

Method

Apparatus.—The Pattern Discrimination Apparatus described previously (2) was used. This device permits pairs or individual series of dot patterns, appearing as pinpoints of light against a homogeneous background, to be presented successively with control of all relevant time intervals. In the present study, *S*s, in groups of eight, sat about 15 ft. in front of the apparatus, viewed successively a series of 64 target patterns, and, in each case, attempted to find each target in the corresponding scope display on the other side of the screen.

Patterns.—The patterns used in this study were part of the series of random patterns employed in a previous study (4). In the present case only the pattern series from one to nine dots was used. The patterns used represented a random sample from the population defined by the limits of the matrix of 112 light points arranged in rows and columns 1 in. apart.

Procedure.—It was necessary to have an objective measure of the recognition or discovery of the target. The scope display, i.e., the target pattern plus the noise pattern, was always presented rotated 36° clockwise or counterclockwise with respect to the same target pattern appearing alone on the opposite side of the screen. The *S*'s task was actually a judgment of whether the scope pattern had been rotated to the "right" (clockwise) or to the "left" (counterclockwise) from the position in which the orientation of the embedded target pattern would be the same as the target pattern alone. Each *S* indicated his response by pressing one of two microswitch buttons located in a covered box on the arm of his chair. These buttons were labeled "R" and "L" to represent the direction of rotation of the scope pattern. The instructions to *S*s indicated that the task was one of finding the target pattern in the scope display and indicating the direction of its rotation. The target and scope patterns appeared simultaneously and were present on the screen for 5 sec. The *S*s were instructed to compare the two displays carefully, and to decide as quickly as possible the direction of rotation. The *S*s were required to make a choice on every trial and were allowed only a single choice. The response, appearing on a control panel, was recorded manually on a prepared data sheet.

Design.—Two main variables were manipulated within the testing period for each group of *S*s: (*a*) the number of dots in the target pattern, and (*b*) the number of noise dots. The target patterns used were composed of dots numbering from two to nine, while the dots in the noise patterns varied from one to eight. For each number of dots, eight different target patterns and eight different noise patterns were used. The assignment of the conditions for each of 64 consecutive trials was based on an 8×8 greco-latin square, with each *S* represented in every cell of this square. In the present case the Greek and Latin letters represented "number of target dots" and "number of noise dots," respectively. The eight occurrences of each symbol represented the eight patterns of a given number of dots. Through the use of the greco-latin square principle, each number of target dots appeared once with each number of noise dots, i.e., the two variables were made orthogonal. Each *S* was presented the sequence of conditions represented by the successive cells of the square taken column by column.

Without further counterbalancing, each unique combination of conditions, for example, a two-dot target pattern embedded in a four-dot noise pattern, would be represented by single patterns only. To obtain a wider sample of pattern combinations, the same greco-latin square was

replicated eight times using different patterns (but the same *number* of target and noise dots) in each cell for each replication (Sets 1–8). The mechanism of this counterbalancing can be visualized in terms of a shift of the specific patterns associated with each column of the square to the adjacent column, with the patterns of the last column shifting to the first in each case. For example, in the cell pairing a two-dot target pattern with a four-dot noise pattern, each of the eight two-dot patterns appeared in turn, paired with a different four-dot pattern in each case. As a consequence of this counterbalancing, each of the 64 cells in the greco-latin square was represented by eight unique combinations of target and noise patterns.

To obtain an even more representative sample of patterns a final kind of replication was employed. The whole experiment, including the eight replications which shifted the patterns in each cell, was performed a total of three times, using in each case a new random set of 64 target and 64 noise patterns (Groups I–III).

Subjects.—The *S*s were 192 basic trainees at Lackland Air Force Base. These men were an unselected sample from the pool of *S*s available to the laboratory for testing. Sixty-four *S*s served in each of the three groups, eight in each of the eight sets. Groups of eight *S*s were assigned in a block to a given experimental condition in order of availability.

Results

The results of the present study are shown in Fig. 1. As a convenience, the target and noise functions have been plotted on the same graph with number of dots on the abscissa and mean percentage of error as the ordinate scale. In interpreting this graph, it should be noted that the data of the three groups and the eight sets within each group have been pooled. Since number of target and number of noise dots are actually orthogonal variables, the mean for each point of the target function was obtained by pooling the data from eight noise-dot conditions and vice versa. Every point on both functions represents data from the same group of 192 *S*s. Mean percentage of error was obtained by summing the error scores for all of the *S*s represented at a given point, dividing by the total possible error (no correction for guessing), and converting this ratio to a percentage. It is apparent from Fig. 1 that a clear substantiation of the experimental hypothesis was obtained. When the number of dots in the target pattern is increased from two to nine, there is a progressive, negatively accelerated drop in errors with a levelling off in the region of eight dots; on the other hand, when the number of dots in the noise pattern is increased from one to eight, there is a progressive increase in errors (apparently linear).

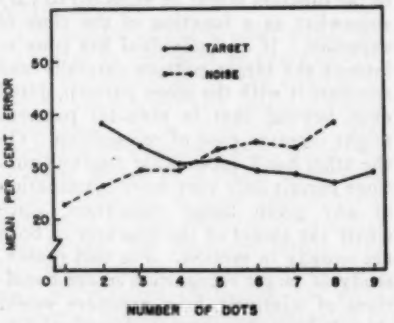

Fig. 1. Mean percentage of error as a function of number of target dots and number of noise dots. The two independent variables were orthogonal; the mean for each point of the target function was obtained by pooling the data from eight noise-dot conditions and vice versa. Every point on both functions represents data from the same group of 192 *S*s.

A somewhat simplified analysis of variance will be described for these data, although a more complete analysis was performed as well. With respect to the primary variables of the study and the significant interactions, the *F* ratios are essentially identical for the two analyses. In the analysis to be described, the data from the three groups were pooled. It will be recalled that the purpose in replicating the experiment was to obtain a wider sampling of patterns. Since each set of patterns was a random sample, the data from each should vary only by chance. An additional pooling procedure was performed for both the simplified and more complete analysis. Since the 64 cells of the greco-latin square represented the 64 consecutive trials of the experiment, each *S* is represented in each cell by only binary data. To avoid the assumptions involved when data of

TABLE 1
SUMMARY OF ANALYSIS OF VARIANCE FOR ERRORS

Source	*df*	Mean Square	*F*
Number of target dots	7	16.83	16.83/1.72 = 9.78*
Number of noise dots	7	30.06	30.06/2.04 = 14.73*
Ss (pooled)	23	5.84	
Targets × noise	49	5.27	5.27/2.21 = 2.38*
Targets × *Ss*	161	1.72	
Noise × *Ss*	161	2.04	
Targets × noise × *Ss*	1127	2.21	
Total	1535		

* Significant at the .01 level.

of this sort are used in an analysis of variance, the scores of the eight *Ss* tested at one time on the same conditions were combined. Note that, as a result of this pooling, the variance between *Ss* is equivalent to the variance between the 24 replications using different pattern combinations (Sets 1–24). With these simplifications, the data may be analyzed in terms of an 8 × 8 matrix with 24 scores per cell (each "score" representing eight binary scores, pooled).

A summary of the analysis is presented in Table 1. Differences between number of target dots, number of noise dots, and the interaction of number of target and noise dots are all significant at the .01 confidence level. The error term for each mean square is indicated in the table (an appropriate error term for *Ss* is not present).

In lieu of separate *t*-test comparisons, a "critical difference" between means at the .05 confidence level was determined. For differences between the means of the target conditions, a standard error term was derived from the target × *S* mean square. For the comparison of target means, a difference between means of 9.31% or greater was found significant at the .05 level. For difference between noise-dot means, using the appropriate error term from the analysis of variance, a difference of 10% was found significant. It is evident from the magnitude of these values that none of the deviations of individual points from the general trends of the curves is significant.

Although the study was not designed to investigate the effect of different target-to-noise ratios, the data could be plotted to yield some information on this point. A plot of mean error as a function of target-to-noise ratios from .5 to 4.5 showed that little further improvement in recognition performance was found beyond a ratio of 3:1.

DISCUSSION

Pattern recognition against a background of noise becomes progressively easier as the number of elements in the pattern is increased. The data suggest, however, that this function is negatively accelerated. Actually, the exact shape of the function might be expected to vary somewhat as a function of the time of exposure. If an individual has time to inspect the target pattern carefully and compare it with the scope pattern, detail even beyond that in nine-dot patterns might increase ease of recognition. On the other hand, most radar display situations permit only very brief examination of any given target indication, since either the target or the operator or both are usually in motion. For this reason, study of target recognition under conditions of relatively brief exposure would seem to have more potential applicability to the operational situation.

The significant interaction between number of target and noise dots should perhaps be discussed briefly. It is recognized that number of dots has a rather peculiar status as an independent variable, since number cannot be varied with-

out varying pattern as well. For this reason it is desirable to use as large a sample of random patterns as possible with each number of dots. Inevitably, however, with a sample of only 24 patterns there may be large and perhaps even significant differences between number of dots which are attributable to pattern variance alone. It is always possible that some statistically significant irregularity in the function for number of dots is due to such associated pattern variance. When subgroups of patterns yield the same function, however, or when the fluctuations of the curve are progressive and consistent with other information, the possibility of an erroneous conclusion is considerably lessened. In the present case, it is apparent that in the pooled data shown in Fig. 1 a large enough sample (192 different scope patterns) is represented at each point to counterbalance most of the pattern variance. In a plot (not included) of the data of the 8×8 matrix, with number of target dots on the abscissa and number of noise dots as a parameter, the individual functions were highly irregular. This irregularity did not appear to be systematic, however, which suggested that the interaction effect was due to random pattern variance rather than a "true" interaction between the primary variables.

The results of the present paper are not meant to represent an exact quantitative determination of the relation of target complexity and noise-to-target recognition. It is doubtful, in fact, whether such a determination could ever be made except in a rough qualitative sense. Although a more sophisticated model such as that provided by information theory, for example, might have been employed, it is a demonstrated fact that all patterns containing the same number of dots are not discriminated or identified with equal ease. Factors such as symmetry and linearity (4), average separation of dots (2), and relative angular orientation (3) must be taken into consideration, to say nothing of the many factors of experience which make some patterns easier to perceive than others. The only practical approach to this problem of obtaining functional relationships with pattern stimuli would seem to be an empirical one: select as wide a sample of representative patterns as possible and trust that most of the pattern variance will be averaged out.

Summary

Recognition of a "target" pattern of dots against a background of "visual noise" was studied as a function of both the number of target dots and the number of noise dots. Data were obtained for each number of target dots from 2–9, using 192 different combinations of random target patterns embedded in random noise patterns varying from 1–8 dots. The *S*s were 192 basic trainees at Lackland Air Force Base.

The results indicate that increasing the complexity of the target pattern by increasing the number of elements improves recognition performance progressively. On the other hand, increasing the complexity of the visual noise produces a progressive decrement in recognition of the target. Increasing the number of elements in the target pattern beyond seven or eight seems to produce little if any further improvement in recognition performance. In general, recognition performance improves as the ratio of number of target dots to number of noise dots increases up to a target-to-noise ratio of about 3:1.

References

1. French, N. R., & Steinberg, J. C. Factors governing the intelligibility of speech sounds. *J. acoust. Soc. Amer.*, 1947, **19**, 90–119.
2. French, R. S. The discrimination of dot patterns as a function of number and average separation of dots. *J. exp. Psychol.*, 1953, **46**, 1–9.
3. French, R. S. The accuracy of discrimination of dot patterns as a function of angular orientation of the stimuli. USAF Hum. Resour. Res. Cent., Res. Bull., 1953, No. 53–3.
4. French, R. S. Identification of dot patterns from memory as a function of complexity. *J. exp. Psychol.*, 1954, **47**, 22–26.
5. Guilford, J. P. (Ed.) *Printed classification tests.* AAF Aviation Psychology Program Research Report No. 5. Washington: U. S. Govt. Print. Off., 1947.

(Received May 20, 1953)

Journal of Experimental Psychology
Vol. 47, No. 1, 1954

EFFECT OF SHIFT IN DISTRIBUTION OF PRACTICE CONDITIONS FOLLOWING INTERPOLATED REST [1]

JACK A. ADAMS AND BRADLEY REYNOLDS

Perceptual and Motor Skills Research Laboratory, Human Resources Research Center

This investigation bears on Hull's inhibition postulates (2, 3) and certain applications of them to human behavior phenomena (5, 6). These postulates specify that an inhibitory state (I_R) is generated each time an organism makes a response. I_R is assumed to have the following properties: (*a*) it develops as an increasing function of number of responses in a sequence and amount of physical work involved in the making of each response, (*b*) it dissipates as an exponential decay function of time, and (*c*) it has drive properties and its dissipation, therefore, is regarded as drive reduction. Since drive reduction is the basis of reinforcing an S–R bond in Hull's theory, habit strength for the ongoing response develops when I_R dissipates. The response reinforced is presumably one of nonactivity, since I_R dissipates under these circumstances. This response has been termed the "resting response" and its associated habit strength, ${}_sI_R$. Thus, the total inhibition ($\dot{I}_R$) developed in responding is comprised of a temporary component (I_R) and a permanent component (${}_sI_R$). $\dot{I}_R$ subtracts from the excitatory potential for the goal response (${}_sE_R$) to yield effective excitatory potential (${}_s\bar{E}_R$).

Kimble (5) and Kimble and Shatel (6) have specified that the development of I_R and ${}_sI_R$ can be inferred from an experimental design where each experimental group is given a different amount of massed practice and then all have a constant rest period. Following rest, the experimental groups resume massed practice. A control group is given uninterrupted, widely distributed practice. Gain over rest for each experimental group is taken to reflect the amount of I_R dissipated, and the functional relationship between amount of I_R and amount of prior massed practice can thus be inferred. The discrepancy between the first postrest trial for each experimental group and the corresponding trial of the control group is taken as a measure of ${}_sI_R$ and permits the determination of ${}_sI_R$ as a function of amount of prior massed practice. This method for measuring ${}_sI_R$ is based on the assumption that negligible I_R develops under the control group condition of widely distributed practice and, as a consequence, little or no ${}_sI_R$ develops. Assuming that the rest period given an experimental group allows dissipation of most of the I_R present, the amount that it is below the control group on the first postrest trial can be taken as an index of the amount of ${}_sI_R$ developed.[2]

In the research to be reported, the foregoing design has been used with the exception that experimental groups given prerest massed training resumed practice after interpolated rest under the widely distributed practice condi-

[1] The experimental work for this study was performed as part of the United States Air Force Human Resources Research and Development Program. The opinions or conclusions contained in this report are those of the authors. They are not to be construed as reflecting the views or indorsement of the Department of the Air Force.

[2] This approach in determining I_R and ${}_sI_R$ functions has been symbolically expressed by Hull (3, pp. 76–77).

tions prevailing for the control group. On the basis of the I_R and ${}_sI_R$ considerations cited, two principal predictions can be made: (*a*) each experimental group should show gain over rest and these gains should be an increasing function of amount of prerest practice (I_R dissipation), and (*b*) the performance of each experimental group during postrest distributed training should stabilize at a level below that of the control group in an amount proportional to the difference between them on the first postrest trial (${}_sI_R$).

Method

Apparatus.—The apparatus employed was a 15-unit Rotary Pursuit Test. This apparatus has been described elsewhere (7). Time-on-target scores were recorded in units of .001 min. on Standard Electric timers. A buzzer sounded and the disks turned for 1.5 sec. before the clocks could be activated.

Subjects.—The *S*s were basic airmen trainees drawn from the population available at Lackland AFB, Texas. They were randomly assigned to five main groups in subgroups of 13–15.

Procedure.—All groups were given 40 trials. Trial length was 15 sec. for all groups. The intertrial rest was 45 sec. for distributed practice and 5 sec. for massed practice. A control group (C) had an uninterrupted practice session under distributed conditions. Experimental groups M-D5, M-D10, M-D15, and M-D20 had initial training of 5, 10, 15, and 20 massed trials, respectively, a 10-min. rest, and the balance of the 40 trials under distributed practice. The *N* in each group was as follows: Group C, 118; M-D5, 112; M-D10, 117; M-D15, 110; M-D20, 108.

Results

To establish the tenability of the hypothesis that these five groups were random samples from the same population prior to differential treatment, Trial 1 scores were subjected to a simple analysis of variance. The *F* ratio was .35, which, for 4 and 560 *df*, did not permit a rejection of the null hypothesis.

Figure 1 shows the performance curves for Group C and the four experimental groups. The break in the curve of each experimental group indicates the point of 10-min. rest. Group C, which had distributed practice throughout the 40 trials, has a curve with an extended curvilinear segment in the early trials and approximate linearity in the later trials. For the experimental groups, massing of practice in the prerest trials has two principal effects: (*a*) performance is markedly depressed below that of performance on corresponding trials of the distributed group, and (*b*) the performance curve has a more abrupt initial curvilinear segment and a proportionately longer linear phase in the later trials than Group C. These general characteristics of psychomotor performance curves as a function of distribution of practice have been noted elsewhere (11).

Each experimental group displays a large gain in mean performance level over the 10-min. rest and the amount of mean gain in each case is given in Table 1. The *t* test for related measures was used to test the null hypothesis for each mean gain, and these values of *t* are also shown in Table 1. All of them are significant beyond the .01 level of confidence. The trend is one of increasing gain as amount of prerest massed practice increases up to 15 trials and then a slight decrease for 20 trials.

Performance on the initial postrest trial for each experimental group is

TABLE 1

Mean Gain over 10-Min. Rest for Each Experimental Group

Group	*N*	Gain in Minutes	*t*	*p*
M-D5	112	.029	9.14	<.01
M-D10	117	.039	10.55	<.01
M-D15	110	.041	12.09	<.01
M-D20	108	.035	8.92	<.01

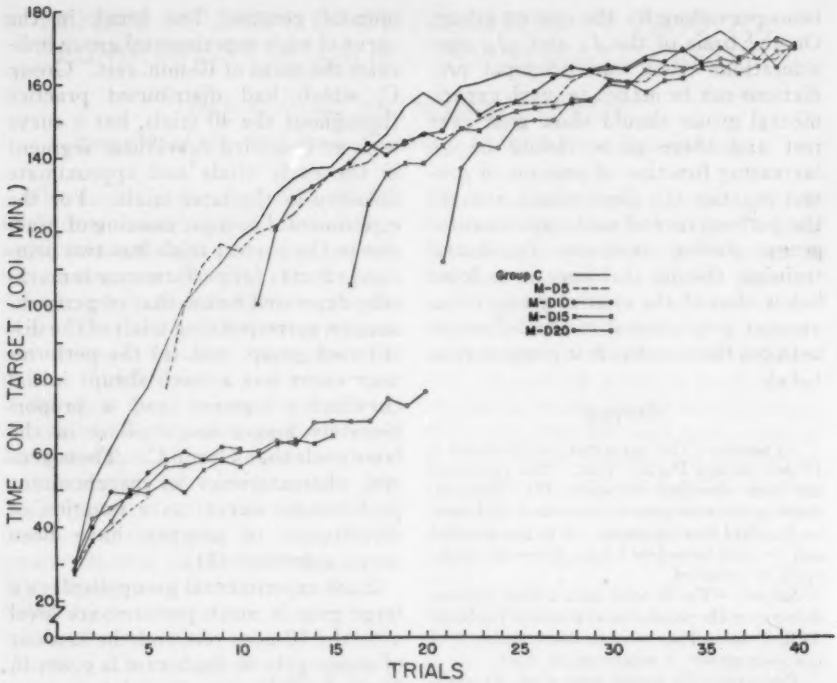

FIG. 1. Performance curves for the control group (C), which had distributed practice throughout training, and the four experimental groups which had massed practice in the prerest trials and distributed practice in the postrest trials

considerably below that attained on the corresponding trial by Group C. This discrepancy, however, readily disappears with practice under the same conditions of distribution as Group C. Figure 1 shows that after rest, each experimental group readily shifts to the level of Group C and that all appear to have about the same rate of shift. This independence of rate of shift and amount of previous massed practice is consistent with previous findings (11). It is noteworthy that all experimental groups are essentially at the same level as Group C in the final trials and show no decremental residual persisting from earlier massed practice. To test the hypothesis that final postrest performance level is independent of amount of prerest massed practice, Trial 40 scores for all groups were subjected to simple analysis of variance. The *F* ratio, for 4 and 560 *df*, was .15, which did not permit the rejection of the null hypothesis.

One possible interpretation of the experimental groups' stabilizing at the same level as Group C is that all groups are approaching the task asymptote, and this serves to mask differences attributable to prerest massed practice. This seems unlikely since all groups have attained a mean level of around .170 min. on Trial 40 which is only 68% of maximum. Another index of performance at the task asymptote is the variance of

TABLE 2

SAMPLE VARIANCE OF MEASURES ON SELECTED TRIALS FOR EACH GROUP

Group	Trial			
	25	30	35	40
C	.0016	.0017	.0015	.0018
M-D5	.0019	.0013	.0017	.0017
M-D10	.0013	.0016	.0014	.0016
M-D15	.0015	.0015	.0015	.0013
M-D20	.0016	.0014	.0014	.0016

scores on the later trials. A marked decrease in variance would be found on these trials if many *S*s had attained a limiting performance level. The variance of measures on several postrest trials for each group is shown in Table 2. There is clearly no tendency for variance to decrease during these trials.

DISCUSSION

First, with respect to I_R. From the operational definition of this construct, it would be anticipated that all experimental groups would display gain over rest and this gain would be some increasing function of number of previous massed trials. Table 1 shows that this prediction is mainly substantiated with the single exception that Group M-D20 shows less reminiscence over rest than Group M-D15. A similar trend of increasing and then decreasing gains over a constant rest period as a function of amount of prerest massed practice is consistent with findings of Ammons (1) and Irion (4). There appears to be no apparent reconciliation of this trend of gains with the I_R construct as currently expressed.

As for ${}_sI_R$, the data do not support the prediction that performance of an experimental group in the postrest distributed trials would stabilize at a level below that of the control group in an amount proportional to the difference between them on the first postrest trial. Instead, all experimental groups readily shifted to the level of the control group in the postrest distributed trials. This finding does not necessarily constitute a clear-cut refutation of the ${}_sI_R$ concept but it does seem to present some difficulty for it. In defense of ${}_sI_R$, it could be contended that practice of the goal response in the postrest distributed trials led to a weakening of the "resting response" by some counterconditioning process and thereby allowed the experimental groups to attain the level of the control group. But since habit interference received little attention by Hull, it is somewhat obscure how his theoretical system could explain such notions, if at all.

One theoretical possibility to account for the present findings is that the patterning of trials under a specified condition of distribution defines a certain level of responsiveness in the organism. This level is some positive function of the intertrial interval. Introduction of a rest period longer than the intertrial rest period being used serves to alter this established level. When practice is resumed after rest, a period of transition is necessary for the organism to attain a level of responsiveness defined by the conditions of distribution prevailing in the postrest session. This period of transition is clearly evident in the present results and shows performance curve characteristics not unlike those attributed to warm-up. A formulation relating the patterning of pertinent stimuli to response level has been advanced by Reynolds (8, 9) for conditioning and has been extended to motor performance by Reynolds and Bilodeau (10).

One implied feature of the analysis should be emphasized. The fact that all experimental groups readily shifted to the level of the control group in the postrest distributed trials suggests that massing of practice leads to a decrement in *performance*, not *learning*. The same performance level for all groups in the final postrest trials would seem to indicate that the same habits have been acquired. This implies that the learning of a psychomotor task is a function of number of trials and independent of conditions of practice distribution. It is recognized, however, that these findings will have to be verified on other types of tasks. It

is possible that massing of practice on some tasks could lead to the acquisition of inappropriate responses which would impair later performance under distributed training conditions.

Summary

Certain implications of Hull's inhibition concepts I_R and sI_R for motor learning were investigated.

All groups were given 40 15-sec. trials with 5-sec. intertrial rest periods for massed practice and 45-sec. rest periods for distributed practice. A control group received distributed training throughout. Four experimental groups had either 5, 10, 15, or 20 trials of massed practice, a 10-min. rest, and the balance of the trials under distributed practice.

Results were as follows:

1. Each experimental group displayed a significant gain over rest. These gains tended to increase and then decrease as a function of amount of prerest massed practice.

2. Performance of the experimental groups in the postrest trials failed to stabilize at a level below that of the control group. Following rest, all experimental groups readily shifted to the level of the control group.

The data were interpreted to imply that massing of practice does not lead to the development of any permanent decremental states. Distribution of practice is regarded as a performance rather than a learning variable.

References

1. Ammons, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393–411.
2. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
3. Hull, C. L. *Essentials of behavior.* New Haven: Yale Univ. Press, 1951.
4. Irion, A. L. Reminiscence in pursuit-rotor learning as a function of length of rest and of amount of pre-rest practice. *J. exp. Psychol.*, 1949, **39**, 492–499.
5. Kimble, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
6. Kimble, G. A., & Shatel, R. B. The relationship between two kinds of inhibition and the amount of practice. *J. exp. Psychol.*, 1952, **44**, 355–359.
7. Melton, A. W. (Ed.) *Apparatus tests.* AAF Aviation Psychology Program Research Report No. 4. Washington, D. C.: U. S. Govt. Print. Off., 1947.
8. Reynolds, B. The acquisition of a trace conditioned response as a function of the magnitude of the stimulus trace. *J. exp. Psychol.*, 1945, **35**, 15–30.
9. Reynolds, B. Extinction of trace conditioned responses as a function of the spacing of trials during the acquisition and extinction series. *J. exp. Psychol.*, 1945, **35**, 81–95.
10. Reynolds, B., & Bilodeau, I. McD. Acquisition and retention of three psychomotor tests as a function of distribution of practice during acquisition. *J. exp. Psychol.*, 1952, **44**, 19–26.
11. Reynolds, B., & Adams, J. A. Effect of distribution and shift in distribution of practice within a single training session. *J. exp. Psychol.*, 1953, **46**, 137–145.

(Received May 15, 1953)

Journal of Experimental Psychology
Vol. 47, No. 1, 1954

THE CONTRIBUTION OF COMPONENT ACTIVITIES TO THE TOTAL PSYCHOMOTOR TASK [1]

EDWARD A. BILODEAU AND INA McD. BILODEAU

Perceptual and Motor Skills Research Laboratory, Human Resources Research Center

Multidimensional psychomotor tracking tasks are often analyzed into component activities. For example, the path of the indicator controlled by *S* can be resolved into the components or dimensions azimuth (*A*) and elevation (*E*). Motion of the control in one dimension displaces the indicator along *A* and motion of the control in another dimension displaces the indicator along *E* (a separate control may even be used for *E* displacements). It is common procedure to measure separately the time the indicator is within prescribed limits or tolerances along *A* and along *E* during a standard practice period. Usually, a third measure is taken as the time the indicator is within the *A* and *E* tolerances simultaneously. Such a measure can be called *AE*, or the total time-on-target index, and is often used as the criterion of response proficiency.

If the relationship among components, and between components and total criterion, were known, considerable use of this knowledge might be made. First, it would be possible to predict accurately criterion proficiency by means of the part proficiencies; second, as will be seen, it might be possible to rearrange the part performances in order to increase criterion performance without increasing the underlying skill at any one of the components. These two accomplishments are possible whether the relationship is positive, negative, or zero. What is necessary is an estimate of the relationship and of its magnitude. Thus, the first step is to ascertain what is already known about these relationships.

General reviews of tracking behavior can be found in Craig and Ellson (2) and in Fitts (4). The relation of *A* and *E* to *AE* was investigated for the Pedestal Sight Manipulation Test (5) by Ellson (3).[2] Ellson reported that *AE* could be predicted quite accurately by proper utilization of the component scores, *A* and *E*. The relationship suggested by his data was

$$A_o \times E_o = AE_p = AE_o \qquad (1)$$

where *A*, *E*, and *AE* represent proportions of time on target, and *o* and *p* identify the proportions as obtained and predicted, respectively. This multiplicative relationship implied the independence of *A* and *E*. In other words, if at any moment *S* is on target in one component he is neither more nor less likely than chance to be responding correctly with respect to the other component.

Actually, on the basis of a *t* test, Ellson rejected the hypothesis that $\overline{AE}_o = \overline{AE}_p$, but since the mean difference was very small he concluded that the error had little actual significance (in a practical sense). The

[1] The experimental work for this study was performed as part of the United States Air Force Human Resources Research and Development Program. The opinions or conclusions contained in this report are those of the authors. They are not to be construed as reflecting the views or indorsement of the Department of the Air Force.

[2] This test, often called the PSMT, presents a moving target on a screen and *S* is required to track the target in *A* and *E* with a gun sight. A third component, range (*R*), requires *S* to adjust a ranging ring to the variable wing span of the moving target plane.

direction of the mean difference indicated a less than chance coincidence of times of scoring in components A and E, in that $\overline{AE}_p > \overline{AE}_o$. Further, the practical independence of tracking (AE) and ranging (R) was demonstrated in like comparisons of $(\overline{AE})\overline{R}_p$ with $\overline{AER}_o$. Statistically, however, the slight negative dependence in times of occurrence of (AE) and (R) was significant ($\overline{AER}_p > \overline{AER}_o$).

Ellson noted that: "The independence of tracking in various dimensions indicated by the data may not represent a general characteristic of multidimensional tracking. The design of the Pedestal Sight, with which this tracking was done, may well prevent the correction of simultaneous errors in several dimensions by a single well-coordinated movement" (3, p. 12).

It is the intent of this paper to show that Ellson's conclusion on independence of the components is more general than at first proposed and to suggest some of the performance (as distinguished from learning) implications of the independent relationship.

The relation between component scores.—The relation between two or more components—A, E, R—under discussion is with respect to times of occurrence of component A and component E scoring and should not be confused with the correlation between arrays of A and E scores.

Thus, an independent (zero) relation refers to a random coincidence of scoring in two or more components, while positive and negative relations refer, respectively, to dependencies leading to greater and less than random coincidences of scoring in the several components. For example, a positive correlation may obtain between arrays of A and E scores, yet the relation between times of occurrence of A and E scoring may be positive, negative, or zero.

As noted previously, the relation between components is indicated by the relation between the obtained and predicted criterion proportions. Component proportions constant, $A = .30$ and $E = .30$, the AE score will vary according to the sign and magnitude of the relation. If the relation is maximally positive, all time in A is obtained simultaneously with time in E. Thus, the AE score must equal .30. (If A and E are unequal, AE is equal to the lesser proportion.) With no relation the two events are independent and thus multiplication of their respective probabilities of occurrence gives the probability of their joint occurrence, i.e., $.30 \times .30 = .09$. If the relation is negative and maximum, however, no time will be accumulated in AE, for $A + E - 1.00 = .30 + .30 - 1.00 = -.40$, a negative number. (If S were more proficient, however, say $A + E > 1.00$, some time in AE would be accumulated.)

At the present status of knowledge concerning tracking behavior, it appears likely that as additional research results become available various tasks will show various relations between components. Since it is believed that at least two tasks already can be assigned, with tolerable error, to the category of independence, the concern of this paper will be largely devoted to making more explicit some of the possible deductions where the relationship approaches zero. Inasmuch as Ellson has dealt only with the prediction of means, a first step is to determine whether obtained and predicted variances show rough agreement within the limits of agreement between obtained and predicted means.[3]

[3] It should be understood that only rough agreement between obtained and predicted means and between obtained and predicted variances is expected, in view of the relatively gross time samples and computational techniques. A number of sources of error is ignored in the following presentation. For example, the use of a fairly long trial allows confusion between the

Analysis of Multidimensional Tracking Data

Prediction of total task variance on the PSMT for two components.—Ellson published a complete table on the tracking performance of 12 well-trained *Ss* on the PSMT, making available the data necessary for computing the variability of the scores of the total (*AE*) task by means of the components *A* and *E*.

The predicted variance of *AE* is the variance of the predicted *AE* proportions. For each *S*, AE_p is computed from $A_o \times E_o$. The variance of this distribution of products can be computed by the equation below.

$$\sigma^2_{A \cdot E} = \overline{A^2 E^2} - \overline{(AE)}^2 \qquad (2)$$

A summary of means and variances based upon Ellson's data and converted to percentage time on target is presented in Table 1. The observed variability of the tracking criterion is 43.49; the predicted variability (AE_p) is 40.40. Thus, the variance has been slightly underestimated. As a first approximation, the agreement between variances is encouraging.

TABLE 1

Mean Percentage Time on Target and Associated Variance for Obtained (*o*) and Predicted (*p*) PSMT Performance Based on Ellson's Data
$N = 12$

Measure	Component		Criterion	
	A_o	E_o	AE_o	AE_p
Mean	69.0	82.6	56.5	57.2
σ^2	20.11	24.54	43.49	40.40

Prediction of total-task means and variances on the PSMT for three components.—The computations summarized in Table 1 above were made with the data of 12 *Ss*. Since the PSMT is acknowledged an unreliable apparatus, it is desirable to corroborate the success of the analysis with additional data from an independent experiment with more *Ss*. An experiment on the PSMT has been previously reported in which the control group practiced in a manner similar to that of Ellson's *Ss* (1).

The 48 *Ss* of the control group were given five successive days of practice at ranging and tracking the simulated target airplane in the conventional manner (5). The *Ss* were also instructed to trigger when they judged themselves to be on target. Each day's practice consisted of 12 trials; eight flights of the plane across the screen (a maximum scoring time of 100 sec.) defined a trial. The 12 trials were administered in three blocks of 4 trials. Blocks were separated by 5-min. rest periods and total testing time each day was about 40 min.

The scoring areas were defined as follows: azimuth, 1 in. (about 17 mils); elevation, 12/16 in. (about 13 mils); range, 5/16 in. (about 5 mils). Calibration was checked both before and after the running of each *S*. Seven different time-on-target measures were taken on each trial from Standard Electric timers. The component measures were the time that the sight was posi-

correlation of the A and B arrays in successive segments of the trial (if the proportions are not constant but change according to the same or different trends, as with learning) and the correlation between times of occurrence of proportions constant throughout the trial. A second source of error is variation in the trial duration, an experimental error; assuming that the independent relationship holds, systematic variations between obtained and predicted means would nonetheless be obtained if the trial time were systematically longer or shorter than assumed. These two sources are examples of errors affecting the mean. These and other factors might tend, however, to have even a greater effect on the variance than on the mean. In addition, using the variance of the products as the predicted variance ignores the expected random deviation of obtained from theoretical proportions. That is, for any *S*, AE_p is the parameter or (theoretical) population mean. If the assumption of independence were to hold, AE_o would equal AE_p in the long run; for any *S*, however, the AE_o and AE_p scores may differ. Essentially, associated with the AE_p score is a variance. Error variance and correlation of errors with the magnitude of the predicted *AE* score are thus ignored when the variance of the products is used to predict the obtained variance.

TABLE 2

MEAN PERCENTAGE TIME ON TARGET AND ASSOCIATED VARIANCE FOR OBTAINED (o) AND PREDICTED (p) PSMT PERFORMANCE BASED ON DATA OF BILODEAU AND MORIN
$N = 48$

Measure	Day 1 Trial 3		Day 5 Trial 3	
	Mean	σ^2	Mean	σ^2
A_o	70.00	60.76	78.49	48.87
E_o	72.45	53.43	78.15	55.10
R_o	29.78	110.38	33.24	120.72
AE_o	54.42	77.06	64.14	87.39
AE_p	51.04	84.17	61.64	94.96
$\overline{AER_o}$	15.86	57.90	20.91	59.02
$\overline{(AE)(R)_p}$	16.37	52.68	21.35	59.89
$\overline{AER_p}$	15.40	49.79	20.51	56.10

tioned within the scoring area in azimuth (A), elevation (E), and range (R) and the total time S triggered (T). Measures were also taken of the criterion combinations: time simultaneously scored in azimuth and elevation (AE); azimuth, elevation, and range (AER); and azimuth, elevation, range, and triggering ($AERT$).

The data for the control group have now been reanalyzed according to the formulae for independent proportions; the results are presented in Table 2 in terms of percentage time on target.

The means and variances of Table 2 have been computed for performance on two different days; Day 1, Trial 3 represents a trial early in practice and Day 5, Trial 3 represents a later trial where *Ss* are more proficient.

Both early and late in practice AE_o is greater than AE_p for each of the 48 *Ss* with a mean difference of about three percentage points. The *SD*'s of the individual *Ss*' discrepancies were 1.30 and 1.25 for early and late practice, respectively. By making the assumption that the relationship is maximum and positive, the mean predicted score is 68.23 (Day 1) and 75.40 (Day 5); by assuming the relationship is maximum and negative, the mean predicted score is 42.67 (Day 1) and 56.64 (Day 5). Thus, the assumption of independence is apparently more appropriate. The direction of the obtained difference indicates a small positive relation between A and E, whereas Ellson reported a negative relation. It may be that the true relation is zero and that the contradiction between these results and Ellson's represents sampling fluctuation in *Ss* or apparatuses. At any rate, assuming the relation to be zero introduces only a very small percentage of error.

The measure labeled $(AE)(R)_p$ represents the criterion predicted as the product of AE_o and R_o. $\overline{(AE)(R)_p}$ overestimates $\overline{AER_o}$ by about .5 percentage point both early and late in practice. The predicted value fell short of the obtained value for 17 and 21 *Ss* on Days 1 and 5, respectively. The *SD*'s of the discrepancies were 1.32 and 1.18. The negative dependence suggested by the overestimation of the criterion $\overline{AER_o}$ is in agreement with Ellson's data; again the difference between obtained and predicted measures is small.

The measure labeled AER_p represents criterion prediction based on the product of the three obtained proportions (A_o, E_o, and R_o). Although the positive relation of A and E noted before is a violation of the assumptions on which the present calculation is based, $\overline{AER_p}$ agrees quite closely with $\overline{AER_o}$. On each day $\overline{AER_p}$ and $\overline{AER_o}$ differ by less than .5 percentage point. This final comparison of obtained and predicted means is of interest chiefly because it indicates that, in spite of the negative relation between AE and R (and presumably between A and R and/or E and R), the positive relation between A and E is sufficient to yield an over-all estimate of positive dependence when the three individual obtained-component proportions are used to predict the criterion.

The variances in Table 2 must be interpreted differently from the means. The variances are treated only with respect to the empirical adequacy of the approximation of the obtained by the predicted variance. For tracking (AE), the predicted variance is greater than the obtained variance both early and late in practice. For (AE)-(R), the predicted variance underestimates the obtained variance on the early trial and slightly overestimates it on the later trial. The variance of the AER_p measure is less than the variance of AER_o on both trials and shows the largest proportional discrepancy.

Certain relationships among variance magnitudes are also suggested. For example, the variance of AE_o is larger than the variance of either A_o or E_o and larger than the variance of AER_o. The variance of AER_o is of roughly the same magnitude as the variances of components A_o and E_o, but smaller than the variances of R_o and AE_o. Component R has the largest variance and, apparently, contributes very little to the variance of AER_o.

If the variances are considered in proportion to the corresponding means, the ratio of the component variance to the component mean is greatest for R_o and about equal for A_o and E_o. The variance to mean ratio for AE_o is roughly twice that for A_o or E_o, while the variance to mean ratio for AER_o is greater than the ratio for AE_o and slightly less than the ratio for R_o. If the reduction in the mean is taken into account, then, the (relative) variance of the criterion is seen to increase with each additional component of the task.

Prediction of total-task means and variances on the Stevens Pursuitmeter for three components.—The Stevens Pursuitmeter is a stick and rudder, compensatory pursuit device with three scoring dimensions (5).

A set of cams drives a pipper about a scope face (deviations in A and E) and an indicator needle about a rudder display dial (R). The R dial is located some 4.5 in. below the scope face, necessitating division of attention between dial and scope. The S's task is to center the pipper by means of the stick (forward-backward motions move the pipper in E, and left-right motions move the pipper in A). The R indicator is centered by means of rudder bars (left bar drives the needle to the left, right bar drives the needle to the right). Perhaps, unlike the pipper of the PSMT, the pursuitmeter pipper can be centered by a single well-coordinated movement of the stick in A and E, and, of course, it should be possible to make the appropriate foot movements (R) at the same time. Time-on-target scores are taken from three component clocks, A, E, and R, and from a total-task or criterion clock, AER. Time on target is accumulated while the pipper and needle are within certain prescribed deviations of center.

In order to test the hypothesis of independence, data were selected without bias from an available data pool, the control group of an experiment conducted by the second author. The Stevens Pursuitmeter as described above was used with a .5-in. (±.25 in.) scoring tolerance for each of the three components. The A and E scoring tolerances were evident to S in a .5-in. square cut-out in the center of a plastic overlay on the scope face. The Ss were instructed to keep the pipper within the square and the R pointer centered at the zero division of the R display. The scoring tolerance of the R component was not indicated.

The Ss were instructed in groups of three, but were tested individually. A total of ten 1-min. trials was administered in two blocks of trials, a training block of six trials and a test block of four trials. The conditions of training and test were the same. Each block began with the first minute of the complete 4-min. cycle of the Pursuitmeter. Within a block the intertrial interval was 30 sec., while the interblock interval was 3 min. The data of one S were discarded, because of clock failure, so that the data presented below are based on 53 cases. The final trial of the test block was arbitrarily selected, without reference to the data, for the present analysis.

Table 3 presents obtained means and sample variances for components and criterion as well as the predicted criterion, AER_p, the latter being ob-

TABLE 3

MEAN PERCENTAGE TIME ON TARGET AND ASSOCIATED VARIANCE FOR OBTAINED (o) AND PREDICTED (p) STEVENS PURSUITMETER PERFORMANCE BASED ON I. BILODEAU'S DATA
$N = 53$

Measure	Component			Criterion	
	A_o	E_o	R_o	AER_o	AER_p
Mean	87.47	76.68	64.79	45.72	44.85
σ^2	124.10	144.82	219.82	251.33	256.13

tained by multiplication of the three individual components. The components are rank ordered A_o, E_o, and R_o in terms of percentage of time on target. AER_o, representing the total task, is considerably below the least of the component measures and the variances are inversely related to the means. $\overline{AER}_p$ compares favorably with $\overline{AER}_o$, the difference being .87. The distribution of discrepancies between obtained and predicted AER scores, taken to two significant digits, showed 27 underestimations of the obtained AER score, 15 overestimations, and 11 cases of zero discrepancy. The SD of the discrepancies was 2.04. The mean differences between obtained AER scores and scores predicted on the assumptions of maximum positive and negative dependencies were, respectively, -17.07 and 13.45, errors considerably greater than the error present when independence is assumed.

A comparison of obtained and predicted variabilities in Table 3 shows fair agreement, for the predicted value is only 2% (relative to the obtained variance) greater than the obtained value.

DISCUSSION

The conclusion on practical independence originally reached by Ellson for the means of two and three dimensions of the PSMT has been verified. Further, the conclusion has been extended to cover three components of the Stevens Pursuitmeter. The present calculations have shown that not only mean criterion performance but also variance can be predicted from component scores with small degrees of error.

To conclude from these data that the relation is essentially zero, for example, is not necessarily to conclude that manipulative corrections between components are always independently performed. It is conceivable that with the present motor tasks corrections were made concomitantly (positive relation) at certain instances, and either independently (zero relation) or successively (negative relation) at certain other instances. If the present computations, based upon trial intervals of 1–2 min., are generally taken to indicate independence, then the safest position is to assume nothing about the relation and determinants of the particular events within the trial, except that over a sizable time interval the events can on the whole be considered to be independent. The present technique at least shows that when the combination score for a trial is used as the criterion of response proficiency, the combination score can be adequately predicted by means of the component scores.

For reasons of chance or otherwise, the events within a single trial for a two-component task can offer the following possibilities for response correction: (a) off-target in one component only, where correcting for A when E is correct is not necessarily the same as correcting for E when A is correct, (b) off-target in two components simultaneously, where one particular patterning of error may lend itself to simultaneous correction but another may not. Evidently, then, a lack of homogeneity in intratrial S–R relations could lead to trial relations of zero. It can be further presumed that a zero relation for a trial is sometimes brought about by the absence of rationale in designing cams and other control devices for presenting the stimulus display.

Since the data from the Pursuitmeter and PSMT have yielded similar conclu-

sions, it is of interest to consider some of the differences between the two tasks. In this way the generality of the finding may best be indicated. The three components of the Stevens Pursuitmeter require compensatory pursuit motions; the PSMT requires simple pursuit motions. Thus, the postulated independence runs across a favorite dichotomy of motor-skills tasks. Ellson, and others, believe that the physical properties of the manipulanda of the PSMT make simultaneous correction of errors in more than one dimension practically impossible. Such a hypothesis is not nearly as tenable with the pursuitmeter. What appears to be true of the Stevens device is that the controls allow simultaneous corrective responses, but *S*s do not, as a rule, make simultaneous corrections in all three components.

The relative contribution of independent component performance to total (criterion) performance.—When the total task is established as the criterion of response proficiency and the nature of the contribution of part proficiencies to the total proficiency is known, there is a sound basis for predicting the level of criterion performance or for calculating what particular redistributions of part proficiencies will do to the level of total proficiency. The present section is devoted to the latter issue, or the question of how the criterion measure changes as scores for the components are rearranged. In order to illustrate the effects of rearrangements among independent components on criterion performance, four examples of arbitrary component values have been selected, and these are discussed below. Later, techniques for effecting rearrangements of component values will be briefly outlined.

Let it be assumed that for three psychomotor dimensions, A, B, and C, $A_o \times B_o \times C_o = ABC_p = ABC_o$. In Part 1 of Fig. 1 are presented hypothetical acquisition curves for the three components. These three linear curves have been chosen to indicate that the components are ranked A, B, and C in terms of scoring difficulty, and that performance improves at equal rates for the three

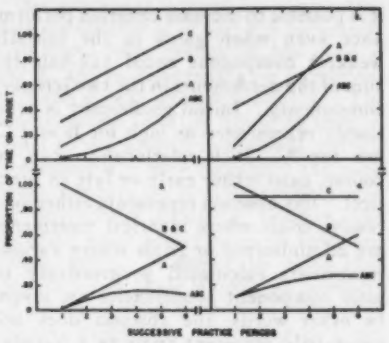

FIG. 1. Each of the four parts of the figure contains hypothetical curves of performance for three components of a tracking task. The curve ABC represents the proportion of time on target in all components simultaneously, assuming independence among the given components.

components. The curve labeled ABC, constructed on the assumption of component independence, shows that the criterion is necessarily (*a*) learned at a positively accelerated rate and (*b*) less than the least of the components (A).

In Part 2 of Fig. 1 the hypothetical component curves have been selected to illustrate the effect on criterion performance of relatively rapid improvement in the (initially) weakest of three components. B and C are represented as equally well performed; A is initially low, but increases more rapidly than B and C and finally surpasses B and C. The curve ABC, computed on the assumption of independence among components, rises slowly at first (from near zero) and then rapidly. However, were the slopes of A and B equal and the slope of C greater than the slopes of A and B, ABC would not rise as rapidly, indicating that the criterion is more sensitive to changes in A (the weakest component). Thus, the implication of the postulated independence among components is to concentrate training attention upon the least well-performed component, whenever there is opportunity for augmenting the scores of that component.

Part 3 of Fig. 1 shows that under the assumption of component independence,

it is possible to increase criterion performance even when gains in the initially weakest component equal one-half the sum of the decrements in the two stronger components. Initial proficiency is arbitrarily represented as high for B and C, low for A. Such relationships can, of course, exist either early or late in practice. The abscissa represents either successive trials where identical treatments are administered or trials where various treatments calculated progressively to alter component proficiencies are given. In other words, the abscissa does not necessarily represent trials as a *learning* variable. Performance on ABC increases at first, provided that A is increased at the sacrifice of proficiency on B and C. It is to be noted that when B and C become relatively small and A relatively large, the maximum for ABC has been passed and ABC decreases.

As an example of the criterion changes suggested in Part 3, suppose *S* can score .90, .90, and .10 or .70, .70, and .30 on a three-component task. In this example it is assumed that *S* is able to make either of the two distributions at will. For the first distribution of part scores the criterion score would be .081, and for the second it would be almost doubled, or .147. In other words, criterion performance would be raised considerably by reducing the two highest components by .20 each and adding .20 to the one remaining component.

However, it may not always be necessary to sacrifice some proficiency on *two* components in order to raise proficiency on the third; deliberate loss of accuracy of the potentially best performed might serve to augment one or even both of the other components. The former possibility is illustrated in Part 4 of Fig. 1, where increases in A are associated with decreases in B while C remains constant. The criterion curve of Part 4 is of the same form as the criterion curve of Part 3, rising gradually to a maximum and falling thereafter with further increases in A.

The four performance situations discussed to this point represent only a sample of possible situations and will probably not cover more than a few of the empirical relationships among components in the research they suggest. Thus, there is little reason to attempt to identify and exhaust the population of possibilities. Suffice it to say that, given differential performance levels at certain multidimensional tasks, a rearrangement among component proportions promises increases in criterion scoring. And increases in criterion scoring seem possible even in situations where the improvement of the lowest component is brought about with considerable loss in one or more of the other components. The augmentation of the criterion might come about without any change in *S*'s general ability level and without additional training on one of the components. Finding the proper techniques may, however, prove to be difficult.

Increasing criterion proportions.—There are two general reasons why one component may lag the others. Obviously, performance on one component may be more difficult than on the others. That is, for example, the path of the target may be more irregular, the rate of target movement may be greater, or the control may be more unresponsive—differential difficulties which would yield differences in proficiencies on the several components practiced separately. In the second general grouping of differential difficulties are those which do not appear under single-component practice or do not change the responses required for true target centering. The following are examples of the latter grouping: (*a*) equal proficiencies might be shown in two components, involving movements of separate controls with the right hand, when the components are practiced separately; if the two controls are presented together, however, proficiency might be much greater on the nearer control, and (*b*) different scoring criteria may be used for the different components, i.e., *A* and *E* may be defined by target widths of 1 in. but *R* may be ½ in. wide. Thus, the proportion of on-target time for *R* would, other things equal, be lower than for *A* and *E*.

Whatever the determinant of the dif-

ferential proficiencies, it has been shown that any increase in score on the low component, even at the expense of proficiency on the easier, will have marked effects on the criterion. The scores of the low component might be increased through additional practice at that component or by instructing *S* to pay relatively more attention to the difficult component as the total task is practiced.[4] Either of these two procedures may perhaps alter criterion performance in a manner best suggested by Part 2 of Fig. 1. The effectiveness of a given procedure in increasing criterion performance may be expected to vary from task to task, according to the nature of the lag in the weak component. If the low score is attributable to a low reaction potential, additional training may be the most efficient way to increase criterion scoring, with attention of secondary or no importance. If single-component proficiency is already at a high level, the attention approach would seem to offer greater promise.

The role of attention appears to be of great importance with both the PSMT and Stevens Pursuitmeter. The protocols from the PSMT indicate that *S*s are reluctant to make the ranging discrimination and the corresponding control adjustments. The relatively low *R* scores suggest that this is an entertainable hypothesis.

On the Stevens device, the rudder (*R*) measure is consistently lower than the *A* and *E* measures whenever the total task is practiced. When, however, components are practiced as components, *R* scores are relatively high. As a component, then, *R* is not inherently difficult, but as a part of the total task, *R* is the most difficult in terms of the same relative score criterion. The case seems to be that for one reason or another attending and responding to *A* and *E* take precedence over attending to *R*.

[4] The word "attention" is used here in its simplest sense. When attending, *S* can be shown to be looking at stimuli of the display and is, presumably, making some discrimination. If *S* is not attending, it cannot be demonstrated that *S* is looking at the display.

One promising method for augmenting *R* at the expense of *A* and *E* is to instruct *S* to spend more time at *R*, less upon *A* and E, and to compensate for deviations in *A* and *E* only when *R* is correct. The augmentation of *AER* would be proportional to *S*'s success in carrying out the directions; if *S* were successful the components would be approximately equal, the dependence maximally positive, and the *AER* score nearly equal to any component score. It should be noted that for such a case there is really no necessity for assuming that *S* increases his skill at any one component, according to a criterion of proficiency at the components tested singly. Instead, it may be assumed that for any trial, *S* is provided with a given reserve of motor skill and is able to attend and respond to fewer signals than the task demands, and that to raise the criterion score *S* must apportion attending and responding so that all components are responded to equally often and simultaneously.

Summary

Data for two multidimensional tracking tasks analyzed by components and total task were presented in an introductory analysis of the relative contribution of components to the total task.

Whether the total task represented two or three components, the analysis (by trials) suggested that, as a first approximation, time of scoring in one component was essentially independent of times of scoring in the remaining components. The hypothesis of independence was tested by multiplication of the separate proportions of time on target and comparing the resulting value with that obtained by observation of the proportion for the total task. The comparison showed that fairly accurate prediction of total-task means and variances can be made, although small but systematic deviations suggest the need for additional research.

The analysis brought out the implication that total-task performance can be improved by a redistribution of the on-target component proportions. Thus, for example, for a two-component task, raising the proportion for one relatively low component at the expense of the second will raise the proportion of the total or criterion. In this way it is suggested that overall proficiency on some tracking tasks might be improved without raising *S*'s general skill level. Methods for bringing about redistributions are still untested.

References

1. Bilodeau, E. A., & Morin, R. E. Proficiency on the Pedestal Sight Manipulation Test with and without the tracking pipper. USAF Hum. Resour. Res. Cent., Res. Bull., 1951, No. 51–27.
2. Craig, D. R., & Ellson, D. G. The design of controls. *Human factors in undersea warfare*. Washington: Committee on Undersea Warfare, NRC, 1949. Pp. 133–151.
3. Ellson, D. G. The independence of tracking in two and three dimensions with the G. E. Pedestal Sight. In G. Finch, Memorandum Report TSEAA-694-2G. AAF Air Materiel Command, Engin. Div., 8 August 1947. Pp. 6–16.
4. Fitts, P. M. Engineering psychology and equipment design. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1287–1340.
5. Melton, A. W. (Ed.) *Apparatus tests*. AAF Aviation Psychology Program Research Report No. 4. Washington, D. C.: U. S. Govt. Print. Off., 1947.

(Received April 27, 1953)

Journal of Experimental Psychology
Vol. 47, No. 1, 1954

POSTREST PERFORMANCE IN MOTOR LEARNING AS A FUNCTION OF PREREST DEGREE OF DISTRIBUTION OF PRACTICE [1]

E. JAMES ARCHER
University of Wisconsin

In a recent paper by Kimble (5) the postrest performance of *Ss* on an inverted alphabet-printing task was studied. It was concluded that the group which had learned under massed practice and had failed to reminisce to a level of performance equal to that of the distributed-practice group was showing the effect of conditioned inhibition. It was Kimble's contention that during massed practice *Ss* developed reactive inhibition and that when this reactive inhibition became superthreshold, they would make a resting response. He further postulated that with this resting response *Ss* developed conditioned inhibition.

Kimble compared the postrest performance of 0-, 5-, 10-, and 15-sec. intertrial rest groups with the performance of a 30-sec. group on its twenty-first trial. Two criticisms can be made of these comparisons. First, the performance of each of the four groups which was compared with the 30-sec. group was performance following a 10-min. rest, whereas the point of comparison for the 30-sec. group was after 30 sec. rest. These comparisons are questionable since the effect of forgetting was not held constant. Another criticism of Kimble's procedure is that he computed four critical ratios before establishing through an analysis of variance that there was any real difference among the five means. The finding that the 0-sec. group apparently had not reminisced to the level of the 30-sec. group could have been a chance phenomenon. In a slightly different context Ammons (1) stated that a massed-practice group after a rest interval will show a permanent decrement as compared to a distributed-practice group. Massed and distributed practice refer of course to prerest conditions.

It was the purpose of this experiment to test for this permanent decrement or conditioned inhibition phenomenon using an inverted alphabet-printing task.

Procedure

Subjects.—The *Ss* in this experiment were introductory psychology students at the University of Wisconsin, and served in groups of about 20 during the first week of the semester when they were naive to the purposes of the task. They were specifically asked not to talk to their friends about what had occurred in the experiment. A total of 243 *Ss* served, and the data for 60 *Ss* in each work condition were randomly selected after discarding papers for various reasons, e.g., failure to follow instructions correctly, task interrupted because of broken pencil. Equal-sized groups were selected to facilitate analysis of the data.

Learning task.—The learning task was inverted alphabet printing as used by Kientzle (4) and Kimble (5) with some modifications. All *Ss* were given booklets containing 24 pages measuring 8½ by 3 in. Fifty-six ¼-in. squares were arranged on each page in a 4 × 14 matrix. The *Ss* were asked to count all pages and number them from 1 through 24 in the upper right-hand corner. They were then instructed as to the nature of the task and told that certain letters of the alphabet were the same whether printed upside down or right side up, these being H, I, N, O, S, X, and Z. This was done to reduce variability as much as possible. A single trial was given on each page.

[1] After this study had been completed it was learned at the May, 1953 meeting of the Midwest. Psychol. Ass. that Schucker, Stevens, and Ellis (7) had performed a similar experiment, the results of which were in press. It is believed that the present study has extended the analyses of data and interpretation.

The *S*s were instructed to print the alphabet as quickly as possible. If they knowingly made any errors, they were to print right over the letter and continue working. At the end of 30 sec. they were told to stop. If they were in the massed-practice condition, they were told to turn the page and continue working. If they were in a distributed-practice group, they waited out their rest period without looking at their papers and on signal turned the page and continued printing the alphabet.

Conditions.—All *S*s received 20 30-sec. trials. One group of 60 *S*s received these 20 trials consecutively without a rest (massed practice). A second group had a 15-sec. rest between each trial. The third group had a 30-sec. rest between each trial. Following the twentieth work trial all groups had a 5-min. rest. Following the 5-min. rest all *S*s had four 30-sec. work trials under massed practice. During the 5-min. rest the distributed groups were instructed that the conditions would be changed somewhat and that they would have four consecutive trials without a rest.

Results

Prerest performance.—As may be seen in Fig. 1 there was considerable facilitation with distributed practice ($F = 26.20$, with 2 and 177 *df*, $p < .001$). Unlike Kimble's results (5), the effect of distribution became apparent very early in learning. Each point on the curve is the mean number of letters printed for 60 *S*s. Except for the early effect of distributed practice, the learning curves do not differ significantly in form from those that could be derived from Kimble's data.

Postrest performance.—The postrest performance of the three groups for Trials 21 to 24 is also presented in Fig. 1. Graphically there appears to be a permanent performance decrement for the massed-practice group in that the three curves show an ordered relationship of superiority of performance as a function of intertrial rest. That these differences however are not significant is indicated in Table 1. An analysis of variance of repeated measures for Trials 21 through 24 indicates the differences among rest groups

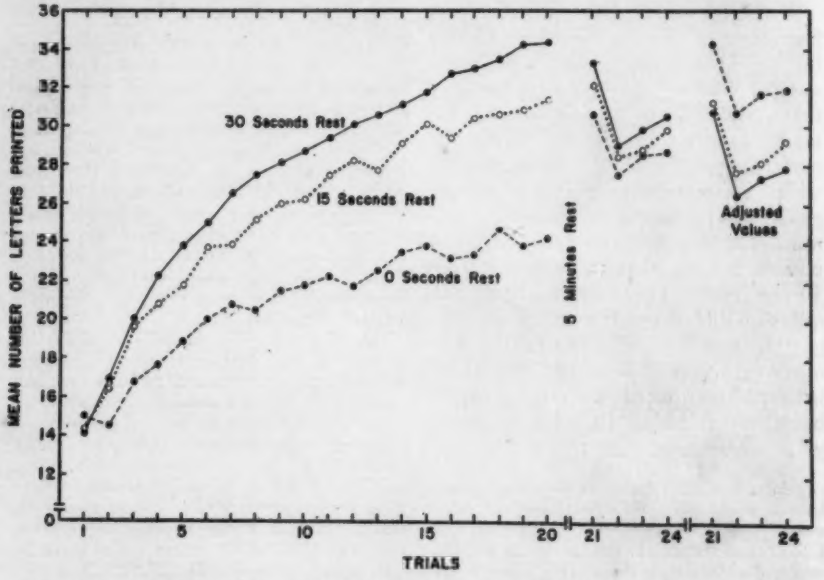

Fig. 1. Mean number of letters printed as a function of practice. The adjusted values for Trials 21–24 were obtained through an analysis of covariance.

TABLE 1

ANALYSIS OF VARIANCE OF POSTREST PERFORMANCE ON TRIALS 21–24

Source	*df*	Mean Square	*F*
Intertrial rest	2	185.35	1.53
Ss/rest groups	177	121.50	18.18**
Trials	3	471.24	70.50**
Rest × Trials	6	15.12	2.26*
Residual	531	6.68	
Total	719		

* Significant at .05 level.
** Significant at .001 level.

are not significant ($F = 1.526$, with 2 and 177 *df*, $p > .10$). The results of this analysis argue against the existence of either conditioned inhibition or a permanent work decrement. There is, however, a significant change in performance as a function of trials, and it is interesting to note that the postrest performance curves bear a close similarity to some of the curves obtained by Ammons (1) on pursuit-rotor performance except that a warm-up decrement is not apparent.

Adjusted postrest performance.—Because of the discrete nature of the learning task an interesting aspect can be examined. In previous learning studies practice usually has been defined as the time during which *S* is allowed to carry out a task. This however is inconsistent with the usual theoretical formulations in which practice is thought of in terms of the *response* rather than the *time* spent in making a response. This notion is analogous to that mentioned by McGeoch (6, p. 455) who pointed out that time was not a parameter of forgetting but rather that what happened during time was the effective variable. The same may also be said of practice. Time is not the proper parameter of practice but rather the responses made during time. If this definition of practice is accepted, then, as may be seen from Fig. 1, the 0-sec. rest group has had less practice than the 30-sec. rest group even though all three groups had 10 min. practice. By the use of analysis of covariance the performance of each of the groups on Trial 21 was adjusted on the basis of number of responses made during the 20 prerest trials. The performance on Trial 22 was adjusted on the basis of the number of responses made on the 21 previous trials. The means for Trials 23 and 24 were also adjusted accordingly. The results of one such analysis are presented in Table 2. Using this procedure the differences between the performance of the three groups are now found to be statistically significant ($F = 10.632$, with 2 and 176 *df*, $p < .001$). In the course of each analysis of covariance a slope constant (b) was determined, and the mean values for each of the three groups were adjusted. The adjusted means for the three groups for Trials 21 through 24 are also presented in Fig. 1 and are identified as adjusted values. (The *F* ratios for Trials 22, 23, and 24, respectively, are 18.898, 23.454, and 17.766; all have 2 and 176 *df*, $p < .001$.) The interesting relationship here is that the curves have now inverted their order and postrest performance of the prerest massed-practice group is now superior to that of the two distributed groups.

TABLE 2

ANALYSIS OF COVARIANCE OF PERFORMANCE ON THE FIRST POSTREST TRIAL (TRIAL 21)

Source	*df*	Mean Square	*F*
Intertrial rest	2	181.07	10.63*
Ss/rest groups	176	17.03	
Total	178		

* Significant at .001 level; $b = .048$, $r = .768$.

Discussion

On the basis of an analysis of variance of unadjusted means following a 5-min. rest on an inverted alphabet-printing task there is no evidence for a permanent work decrement or conditioned inhibition. The postrest performance curves, except for the absence of a permanent work decrement and a warm-up decrement, are somewhat as would be predicted from Ammons' (1) theoretical formulations. There may be several reasons for the absence of the warm-up decrement: (*a*) the loss of perceptual motor sets in an inverted alphabet-printing task may not be of the same magnitude as in a pursuit-rotor task; (*b*) the period of time during which a warm-up decrement may be observed in inverted alphabet printing may be considerably less than 30 sec. and therefore was not detected in the first postrest trial. The very marked drop in performance on Trial 22 (second postrest trial) may be due to a motivational factor, e.g., *S* adjusting his work output to the expected time he will have to work, or as Ammons proposes, a redevelopment of a temporary work decrement.

If a response, rather than the time during which a response is made, is accepted as a unit of practice, then adjustment of the postrest means on the basis of prerest responses seems justifiable.

A theoretical explanation of the performance superiority of the massed-practice group, in terms of adjusted values, over the distributed-practice groups is of necessity quite conjectural since the mean values were adjusted on the basis of this redefinition of practice. One possible explanation of this paradoxical superiority is as follows. In Hull's (3) theoretical formulations reactive inhibition is a negative drive. In the present experiment it would also be an irrelevant drive. The relevant drive was the motivation to complete the task to *E*'s satisfaction. Along the lines of Amsel (2) and Spence and Taylor (8) it would seem reasonable that the irrelevant drive of reactive inhibition and the relevant drive, whatever its specific name may be, could summate to produce the total effective drive state. In Hull's formulation the excitatory potential is equal to a product of habit strength and drive. Furthermore, habit strength is said to be proportional to number of reinforcements. In the present experiment it seems reasonable to define a reinforcement as a unit of practice which would be printing one letter of the alphabet. If these are correct interpretations of Hull's theory, then the results of the adjusted means are as expected. Habit strength, now proportional to number of letters printed, has been equated through an analysis of covariance, and the different levels of performance for the adjusted mean values of Trials 21 to 24 are a reflection of different levels of excitatory potential as a function of different degrees of drive. This is, of course, making the rather tenuous assumption that the negative drive state of reactive inhibition can summate with the positive drive state to give an over-all composite drive. Crudely, the results suggest that during massed practice *S* finds each response more effortful but he learns more per response.

Summary

The results of this experiment may be summarized as follows:

1. As in previous experiments distributed practice facilitated the performance of an inverted alphabet-printing task.

2. After a 5-min. rest the massed-practice group showed a significant amount of reminiscence.

3. The postrest performance of the massed and distributed groups was not significantly unequal. There was no evidence of a permanent work decrement. There was, however, a significant trial-to-trial effect with a marked decrease in performance from the first to the second postrest trial and thereafter a slow increase in performance.

4. When practice was redefined as the making of a response rather than the time during which a response is made, the four sets of three means of postrest performance were adjusted on the basis of the number of previously made responses. This adjustment was done through a series of analyses of covariance and the differences between the means were statistically significant. The adjusted performance of the massed group was superior to that of the distributed groups for the four postrest trials.

5. A possible explanation of this postrest superiority of massed practice, in terms of adjusted means, over distributed practice was offered.

References

1. Ammons, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
2. Amsel, A. The combination of primary appetitional need with primary and secondary emotionally derived needs. *J. exp. Psychol.*, 1950, **40**, 1–14.
3. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. Kientzle, M. J. Properties of learning curves under varied distributions of practice. *J. exp. Psychol.*, 1946, **36**, 187–211.
5. Kimble, G. A. Performance and reminiscence in motor learning as a function of the degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500–510.
6. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.
7. Schucker, R. E., Stevens, L. B., & Ellis, D. S. A retest for conditioned inhibition in the alphabet-printing task. *J. exp. Psychol.*, 1953, **46**, 97–102.
8. Spence, K. W., & Taylor, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

(Received May 25, 1953)

Journal of Experimental Psychology
Vol. 47, No. 1, 1954

PRIMARY STIMULUS GENERALIZATION OF THE GSR UNDER TWO CONDITIONS

DELOS D. WICKENS, HAROLD M. SCHRODER, AND JOHN D. SNIDE

The Ohio State University[1]

Although some form of primary stimulus generalization gradient is employed in nearly all psychological systems, both the exact empirical form of the gradient and its theoretical basis have been questioned. Lashley and Wade (9) have challenged the interpretation of primary stimulus generalization as it was employed by Hull (5, 7). In addition Razran and others (10, 11) have raised the question as to whether or not the gradient that Hull used was empirically correct. This gradient is concave upward and was obtained from an experiment by Hovland (3). The purpose of the first experiment in this paper was to repeat Hovland's experiment in a modified form with a larger number of *S*s; and the purpose of the second experiment was to introduce a procedural variation (a form of partial reinforcement) which might throw some light upon the theoretical nature of stimulus generalization.

METHOD

Apparatus.—The apparatus consisted of three Jackson oscillators which generated the conditioned stimuli, a power source for producing the shock, a General Electric Photoelectric Recorder and an RCA Volt Ohmyst for recording the GSR, and an electronic timing unit which controlled the onset and duration of the conditioned and unconditioned stimuli. The GSR circuit employed was that described by Haggard and Gerbrands (2).

Procedure.—In Exp. 1 *S* was seated in a soundproofed room and told that he was serving in a conditioning experiment and that he would hear a tone in the earphones which would be followed by a shock, but there was no particular response that he needed to make. After adjustment of electrodes and earphones *E* left the room, and with the door open presented two shocks alone. A DC current was employed with 120,000-ohms resistance in series with *S*. The first of these shocks was fairly weak, and the last was stepped up to 150 v., the level which was used throughout the experiment for all *S*s. The *S*s were then given one tone alone, and following this the door between *E* and *S* was closed and the experiment proper began.

Sixteen reinforced trials on a single tone were administered during the training. The interstimulus interval varied from 20 to 60 sec. with a mean interval of 45 sec. The extinction procedure of eight nonreinforced trials followed immediately with no interruption. The average interstimulus intervals remained the same, but were counterbalanced from *S* to *S* so that the group average was the same from one trial to the next. Any given *S* was extinguished to only one of the three tones employed, and this is the major difference between the present experiment and Hovland's, for in his experiment each of the *S*s was tested to each tone in a counterbalanced order.

The frequency of the three tones employed was 153, 468, and 1000 cps. These were frequencies used by Hovland and are presumably separated by 25 j.n.d. Each tone was 40 db above threshold as determined by the judgments of ten *S*s.

Counterbalancing was employed to equalize the frequency of the tone employed in the extinction series and to equalize the frequency with which the extinction tone was higher or lower than the training tone. Hereafter the group trained and tested on the same tone will be called Group 0, the group with 25 j.n.d. between training and test tone Group 1, and the 50-j.n.d. group will be called Group 2. There were 24 *S*s in each of these groups.

In Exp. 2 the procedure was identical with Exp. 1 save that 12 times during the training series a click was sounded in the earphones. This click was interspersed in a prearranged random order among the tones and it was never reinforced. The loudness of the click roughly matched that of the tone. There were 24 *S*s in each of the three groups of this experiment.

[1] This research was supported in part by a grant-in-aid from funds allocated by the Ohio State University Research Foundation to the University for support of fundamental research.

Subjects.—The 144 Ss of this experiment were students in the elementary psychology classes of The Ohio State University.

Results

Experiment 1.—The major results for Exp. 1 are presented in Fig. 1. The small numbers, 0, 1, and 2, represent the mean score for the tone in question at the various extinction points. The data are smoothed by a method of fitting polynomials, and the obtained polynomials are presented in the upper part of the figure.

These equations were tested for the significance of the difference of their displacement term and their slope term by *t* test as described in Kendall (8, Vol. 2, pp. 141–167).

The y_0 and the y_1 curves differ in displacement from the y_2 at the 1% level, but they do not differ significantly from each other. The findings for the slope terms are similar save that the difference between y_1 and y_2 is at the 5% level, while the difference between y_0 and y_2 remains at the 1% level.

In Fig. 2 are presented generalization gradients computed from the fitted curves for the first, second, fourth, and eighth extinction trials.

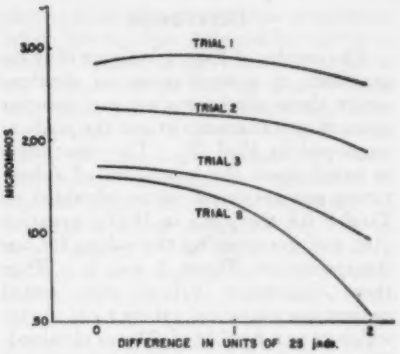

Fig. 2. Gradients of generalization after different numbers of extinction trials

It is apparent that the curves become progressively steeper with increased extinction. It should be noted that this theoretical gradient on the first trial is at variance with the empirical data where both of the generalization tones produce responses of greater magnitude than does the training tone.

Experiment 2.—The extinction curves for Exp. 2 are presented in Fig. 3. There are no significant differences among these three curves, and they do not present any consistent gradient.

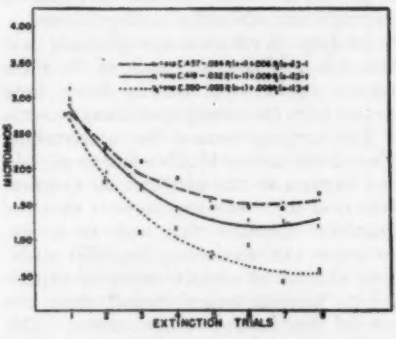

Fig. 1. Extinction curves for the training and generalized tones in Exp. 1

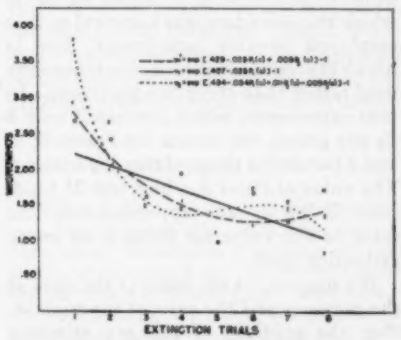

Fig. 3. Extinction curves for the training and generalized tones in Exp. 2

Discussion

The results of Exp. 1 indicate that the gradients of generalization as obtained under these conditions are not concave upward as is characteristic of the gradient employed by Hull (7). This conclusion is based upon the procedure of substituting our empirical values obtained on Trial 8 for the ${}_{s_1}H_R$ in Hull's equation (19) and determining the values for our Generalization Tones 1 and 2. When these theoretical values were tested against our empirical values a chi square which gave a p value of .02 was obtained. Thus our data do not fit his predicted concave upward gradient.

When a straight line was fitted to our data for Trial 8 and the empirical results tested against the theoretical predictions, it was not possible to reject the hypothesis that the gradient could be described by a straight line. However, when the obtained theoretical curve was extrapolated to a point 25 j.n.d. from our Test Tone 2, the theoretical value obtained was slightly negative. In other words such a fit as this would imply that generalization effects operate on only a rather narrow range of the dimension in question. Thus, although there are no statistical grounds for rejecting a straight line as describing the generalization gradient, this gradient does not seem to be acceptable on psychological grounds.

The findings of a previous but unpublished experiment by the senior author in which the procedure was identical to that employed in this experiment, but in which Hovland's four stimulus tones were used rather than three, are pertinent. In that experiment, which contained only 8 *S*s per group, the trends for Points 0, 1, and 2 paralleled those of this experiment. The value of Point 3, which was 25 j.n.d. from Point 2, was approximately the same as the value for Point 2 on every extinction trial.

We suggest, on the basis of the data of the previous and the present experiment, that the gradient of primary stimulus generalization is actually bell-shaped, and that its steepness will vary as a function of the number of extinction trials which have been administered. It should be noted incidentally that qûite different predictions in certain types of experiments involving discrimination would be given if this gradient rather than the concave upward gradient were employed.

The steepening of the gradient as a result of successive extinction tests is not an atypical finding (1, 3). The data of the present experiment do not give any evidence for the gradient on the first test trial, although the fitted curves do produce a slight gradient. It is possible that the empirical results differ from the theoretical curves because there is some form of specific adaptation to the training tone as a consequence of its repeated presentation. This may depress the response to the training tone relative to the generalization tones and thus suppress the gradient on the first trial.

Regardless of whether or not a gradient is present on the first test, the experiment suggests that at least the manifest generalization gradient is not a constant but rather a function which varies with different parameters. One such parameter is the number of preceding nonreinforcement trials. Other parameters for which empirical data are available are the amount of training on the original stimulus (4) and the level of motivation of *S* (11). It is apparent that such a view of generalization increases the difficulties involved in utilizing this concept in our miniature and quantitative systems, but if correct it should increase our precision of prediction when this concept is used.

In Exp. 2 no clear-cut gradient was found in any trial and each of the three curves could very readily have been drawn from the same population of scores.

The original reason for undertaking Exp. 2 was to test Hull's effort to explain the flatness of the gradient on the first test trial (6). His argument is that the gradient becomes flattened as it approaches the maximum, and this situation obtains in certain types of experiments because other stimuli than the crucial discriminative one carries habit loadings which must be extinguished before the gradient can be disclosed. Apparatus cues are examples of these

other stimuli. The addition of these stimuli raises the ${}_{S}E_{R}$ level and in this sense increases resistance to extinction. Actually the nonreinforced click was introduced in Exp. 2 in an effort to test his hypothesis. Briefly it was assumed that some reaction might be given to the generalized stimuli simply because they were auditory events like the training stimulus, and by extinguishing this tendency the tonality gradient could be demonstrated. Obviously no such result was obtained. Assuming there is an analogy between apparatus cues and the function of the click, this experiment does not support Hull's analysis.

If the occurrence of the nonreinforced click during training may be thought of as producing a partial reinforcement situation, then several possible explanations of the results are available.

One such hypothesis is essentially an empirical one. The result of Exp. 1 and others as well (1, 3) have demonstrated that the gradient of generalization is a product of, or at least made manifest by, the process of extinction. Therefore, one might postulate that any condition—and partial reinforcement is undoubtedly one of these—which increases resistance to extinction will tend to flatten the generalization gradient. There is some internal evidence from our two experiments in support of this statement. A t test conducted on the data of the eighth trial between the responses to Tone 2 for the control and the experimental groups gives $.05 > p > .01$. Thus it might be said that the failure to obtain a generalization gradient in Exp. 2 arose from the fact that the resistance to extinction in the Tone 2 group was high.

Another hypothesis may be found in Hull's explanation of the manner in which partial reinforcement operates to produce stimulus-response connections which are especially resistant to extinction (7, p. 120). He assumes that in partial reinforcement training traces of previous nonreinforced trials become a part of the molar trace which is reinforced on the subsequent trial. Therefore the stimulus which elicits the response in a partial reinforcement situation is one which includes the traces of previous stimulation which has not been followed by reinforcement. Hence, there is less difference in the stimulus trace during reinforcement and extinction for a partially reinforced group than for a group trained under 100% reinforcement. Essentially, then, the greater resistance to extinction for partial reinforcement is explained in terms of stimulus generalization. Therefore, although the identical physical stimuli were employed in Exp. 1 and 2, the actual differences in similarity of the afferent molar traces would not be the same, and the stimuli operating in Exp. 2 would be considered to have been drawn from a narrower range of the dimension. Obviously, then, one would expect their ordinate values to be more alike. Thus Hull would seem to need to predict that partial reinforcement would produce an empirical flattening of the gradient. In terms of his psychological variables, however, this empirical flattening is not the result of shifts in the ordinate values, but rather a change in the abscissa values. If the use of the click does produce a partial reinforcement type of situation, then it could be said that the results of Exp. 2 offer some support for Hull's interpretation of partial reinforcement.

It is obvious that the results of these experiments may also be given a perceptual type of explanation. Such an interpretation could make much of Exp. 2. It could state that the training is such as to cause the subject to perceive the relationship between tone and shock, and click and nonshock. This perception would lead to high responsiveness to any stimulus with the tonal characteristic.

Summary

In Exp. 1, 72 *S*s were given 16 pairings of tone and shock, and were then extinguished to one of three tones. These tones were either the same one on which training was given or tones which differed from it by 25 or 50 j.n.d. There were 24 *S*s in each extinction group, and each *S* was extinguished to only one tone. In Exp. 2, 12 nonreinforced clicks were presented during the training, but in all other aspects the experiment was identical with Exp. 1.

It was concluded that the generalization gradient as found in Exp. 1 is bell-shaped rather than concave upward and that it is generated by the process of extinction.

In Exp. 2 no gradient was found. This fact was interpreted in terms of the concept of partial reinforcement.

It is suggested that at least the manifest gradient of generalization is not a constant function, but a function which varies with a number of parameters, one of which is the number of nonreinforced trials.

References

1. Brown, J. S. The generalization of approach responses as a function of stimulus intensity and the strength of motivation. *J. comp. Psychol.*, 1943, **33**, 209–226.
2. Haggard, E. A., & Gerbrands, R. An apparatus for the measurement of continuous changes in palmar skin resistance. *J. exp. Psychol.*, 1947, **37**, 92–98.
3. Hovland, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 279–291.
4. Hovland, C. I. The generalization of conditioned responses: IV. The effects of varying amounts of reinforcement upon the degree of generalization of conditioned response. *J. exp. Psychol.*, 1937, **21**, 261–276.
5. Hull, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
6. Hull, C. L. Simple qualitative discrimination learning. *Psychol. Rev.*, 1950, **57**, 303–313.
7. Hull, C. L. *A behavior system*. New Haven: Yale Univ. Press, 1952.
8. Kendall, M. G. *The advanced theory of statistics*. (2nd Ed.) London: Charles Griffin, 1948.
9. Lashley, K. S., & Wade, M. Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.
10. Littman, R. A. Conditioned generalization of the galvanic skin response to tones. *J. exp. Psychol.*, 1949, **39**, 868–882.
11. Razran, G. H. S. Stimulus generalization of conditioned response. *Psychol. Bull.*, 1949, **46**, 337–365.
12. Rosenbaum, G. Stimulus generalization as a function of level of experimentally induced anxiety. *J. exp. Psychol.*, 1953, **45**, 35–43.

(Received May 11, 1953)

Journal of Experimental Psychology
Vol. 47, No. 1, 1954

INDUCTION EFFECTS IN ESTIMATES OF EXTENT

T. OBONAI

Tokyo University of Education (Tokyo Bunrika University)

Numerous facts indicate that the perception resulting from stimulation of a retinal area is influenced by the states of surrounding areas. For example, (*a*) a patch of gray or of color tends to be tinged with a color complementary to an adjoining one, and (*b*) a colorless surface appears either brighter or darker depending on whether it is brighter or darker than a surrounding area. Köhler considers a field theory of perception in the following terms: ". . . a theory of perception must be a field theory. By this we mean that the neural functions and processes with which the perceptual facts are associated in each case are located in a continuous medium; and that the events in one part of this medium influence the events in other regions in a way that depends directly on the properties of both in their relation to each other . . . The field process may be responsible for any influence which the process exerts upon other percept process" (**3**, pp. 55–66).

Phenomena of the sort discussed by Köhler may be included under the term "spatial induction" (**2, 7**), and in fact, for at least some of its testable implications, Köhler's view is difficult to distinguish from Hering's (**2**). For example, according to Hering, a white process induces a black process in its whole surrounding, the strength of induction decreasing with distance. In the present treatment induction effects are considered to be additive, i.e., amounts of induction are assumed to be representable as algebraic sums of partial inductions. Such an assumption would probably be denied by Köhler.

We consider, first, the perception of a line or a point. What effect can a line or a point have upon a surrounding area? How does induction vary when a line changes its length? To answer this question the following experiments were carried out.

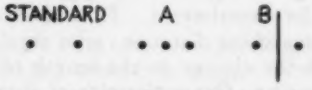

FIG. 1. Stimulus figure with dot or line

In Fig. 1, the distances between the outermost points of A and B, although actually equal, appear to be unequal in length, owing to the different inducing effects of the interpolated inner stimulus. Using figures of this sort, I have attempted to determine how the inducing effect varies with the size of an inner interpolated stimulus that varies from zero to a considerable length.

In making the measurement, a stimulus configuration was observed at the distance which just allowed the outer dots to remain visible. (Each dot was .5 mm. wide and 1 mm. long. The *S* usually sat at a distance of 5 m. from the object.) The procedure employed in our experiments, conducted by the method of constant stimuli, does not need special description. The distance between the standard stimulus dots shown in Fig. 1 was varied to match the distance between the outer dots of A and B. The *S*s were allowed to use the terms "greater," "uncertain or equal," and "smaller," with reference to the test diagram. The frequency of each answer was determined and recorded for each value of the variable used.

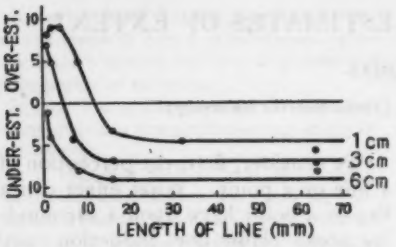

FIG. 2. Relation of under- or overestimation to length of line

Figure 2 presents a typical result of the experiments. It is seen that the matching distance varies regularly with the change in the length of the inner line. Overestimation of distance between dots decreases with the increase in the size of the inner line, and finally becomes an underestimation that increases as the line becomes longer. The change takes place slowly at first, then rapidly, and finally slowly again. Many factors govern the course of the curves. In the first place, the position of a given curve depends upon the distance between the outer two dots: as the distance between the dots is increased, the position of the curve becomes lower.

A similar effect is shown by a figure in which a standard line is extended in the manner of Fig. 3 (after Lewis, 4).

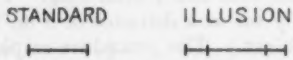

FIG. 3. Stimulus figure with line added

Lewis concerned himself chiefly with the problem of the Müller-Lyer illusion, and made observations with various modified forms of this illusion. He regarded a diagram in Fig. 3 as the limiting case of the Müller-Lyer illusion in which the added segments make an angle of zero degrees with the direction of the central line. The amount of illusion was measured in diagrams with terminal parts of the

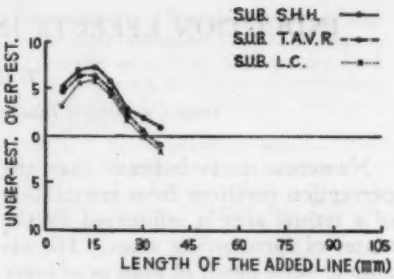

FIG. 4. Relation of under- or overestimation to length of added line

various lengths 5, 10, 15, . . . 35 mm. The quantitative result is shown in Fig. 4.

Lewis's experiment gives a result which is in agreement with the result of an experiment that I have performed, using the diagram shown in Fig. 5. In general, the inner distance

FIG. 5. Stimulus figure with outer lines

of the right-hand figure seems much larger than the comparable distance of the left-hand figure. In the figures used as stimuli the distance between the inner line was 2 cm.; the size of interval between the inner lines and the outer lines was varied by varying outer lines only. These diagrams were observed at a distance of 2 m. By

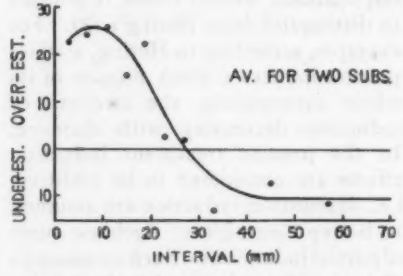

FIG. 6. Relation of under- or overestimation to spatial interval

referring to Fig. 6 it will be seen that the influence of the outer lines upon the inner space varies considerably with changes in the spatial interval between the two.

The question arises: do similar effects occur in other types of configurations? To answer this question, I have conducted measurements with many designs of various sorts. A number of "illusion" designs are shown in Fig. 7.

Designs A, B, C, D, etc., on the left in Fig. 7 indicate that the distances between points in the illusion seem much greater than the comparable distances in the standard. In Design I, when the two circles are put together

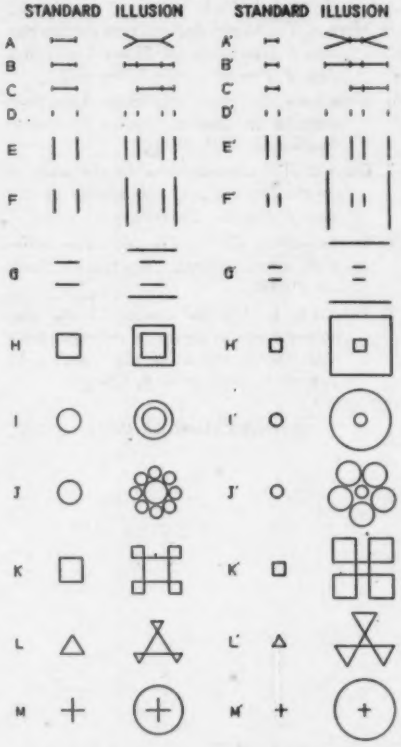

FIG. 7. Illusions used as stimulus figures

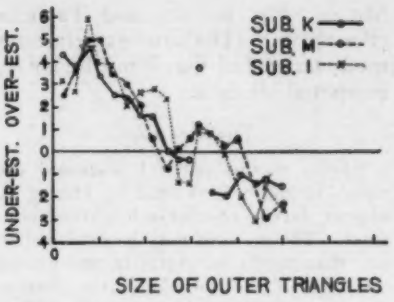

FIG. 8. Relation of under- or overestimation to surrounding figures

to form a ring, the inner circle seems larger than when it is alone. Similar effects are obtained with the other designs.

The designs on the right in Fig. 7, on the contrary, show the opposite induction. In Design A' the middle portion of the illusion design appears shorter than the comparable length in the normal design. Other designs on the right in Fig. 7 show a similar effect.

In the examples cited we see that lines, circles, squares, triangles, etc. seem larger when they are adjacent to small-to-medium extents (the confluence effect) and smaller when they are adjacent to larger extents (contrast effect).

The general result is supported by the earlier data of Morinaga (6),

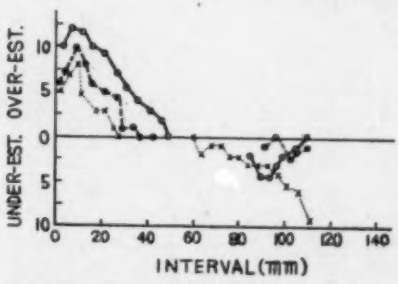

FIG. 9. Relation of under- or overestimation to surrounding figures (based on Tamaike's data [9])

Mitsui (Fig. 8) (5), and Tamaike (Fig. 9) (9). (The latter experimenter used Design I of Fig. 7 and I have recomputed his data.)

Discussion

Earlier in this paper I discussed the term "induction" as used by Hering to indicate factors of color or brightness contrast. The concept may be extended to one that might be useful in accounting for some of the findings of the present experiment. Such an account is different from one given, for example, by Bernstein (1), for it supposes not only that the phenomenon of induction may be due to physiological irradiation in the cerebral cortex (as well as in the retinal layers and thalamus), but also that induction effects must be represented as resultants of antagonistic processes. According to this view, a point or a short line will stimulate a small area of the retina, and hence, since small parts of the projection areas are excited, the amount of spread is also small. Conversely, a long line provides a large degree of spread. Presumably, the excitatory process initiated at a definite cortical point diminishes in intensity with distance from that point. The present account presumes that overestimation and underestimation of visual extents represent antagonistic processes. In any case, it has been shown that extents judged in the presence of small-to-medium extents are overestimated. The same extents, judged in the presence of large extents, are underestimated.

Summary

A series of observations, with many figures, has shown that the sizes of lines, circles, squares, triangles, etc., are (*a*) overestimated when they appear near small-to-medium extents and (*b*) underestimated when they appear near large extents. The data are considered in relation to a concept of physiological induction.

References

1. Bernstein, J. *Lehrbuch der Physiologie.* 1894.
2. Hering, E., Grundzüge der Lehre vom Lichtsinn. In Graefe-Saemischs, *Handb. d. ges. Augenhlk.* (2nd Ed.), 1905.
3. Köhler, W. *Dynamics in psychology.* New York: Liveright, 1940.
4. Lewis, E. D. Confluxion and contrast effects in the Müller-Lyer illusion. *Brit. J. Psychol.*, 1909, **3**, 21.
5. Mitsui, T. Minor studies from the psychological laboratory of Hosei University. *Jap. J. Psychol.*, 1927, **2**, 856–872.
6. Morinaga, S. Some conditions of contrast-assimilation illusion. *Saikin Shinrigaku Ronbunshu*, 1935, 28–48.
7. Obonai, T. Contributions to the study of psycho-physiological induction (1–41). *Jap. J. Psychol.*, 1933–1953.
8. Sherrington, C. S. *The integrative action of the nervous system.* New York. Scribner's, 1906.
9. Tamaike, J. On the change of the phenomenal size of figure in correspondence with the structure of the visual field. *Jap. J. Psychol.*, 1933, **8**, 579–587.

(Received March 27, 1953)